School of Philosophy, Psychology and Language Sciences

A Contrastive Study of Reporting in Master's Theses in Native Chinese and in Native English

By Xiaoli Zhang

Dissertation Presented for the Degree of MSc

in Applied Linguistics

The University of Edinburgh

August 2008

Contents

Acknowledgement

I owe innumerable thanks to people who give me support and assistance in completing this project.

The first goes to my supervisor Hugh Trappes-Lomax, whose considerable encouragement and valuable feedback kept me from losing confidence when I got frustrated with my dissertation. My project could not have been completed without his great patience and sustained guidance.

I'm deeply grateful to my family who always give me warmth and support.

Thanks also to my friends who helped me a lot with my data collection.

Thanks to all.

This work is dedicated to my parents and my beloved Kai.

Statement

I hereby declare that this dissertation was of my own composition except where proper use of quotes and references were indicated, and that this dissertation has not been submitted for the award of any other degree.

Xiaoli Zhang

Abstract

Previous studies mainly focus on reporting practices of English-language writing either by native-speakers-of-English or second-language learners of English. This study investigates similarities and differences of reporting between Chinese theses by native-speakers-of-Chinese and English theses by native-speakers-of-English from four aspects: integral-ness & prominece, reporting forms, categories and tense of reporting verbs, and functions of reporting. 80 theses for master's degree are collected from 8 disciplines, among which 40 are Chinese theses and 40 are English theses. Generally speaking, English writers use more reporting statements than Chinese writers, including integral citations and non-integral citations. Summary and generalization are used most frequently in both Chinese data and English data in five reporting forms. English writers use more reporting statements in each reporting form. Textual and research verbs are employed more frequently in English corpus than in Chinese corpus, but mental verbs occur in Chinese corpus more frequently. Tense of reporting verbs in Chinese is more completed than those in English. Three reporting functions are identified in Chinese and English data, among which background is used most frequently while support is least used. Findings of this study can be used as a basis for investigating why Chinese learners of English use reporting language differently compared with native speaker of English, and also can shed light on pedagogical implication of teaching academic writing to Chinese learners of English.

Chapter 1 Introduction

1.1 Background and rationale of this study

Reporting is an important linguistic phenomenon in academic writing. Writers introduce background of present research and review previous studies to 'create a research space' (Swales 1990). Making reference to claims, activities and findings of previous researchers and evaluate them appropriately is crucial in constructing a research paper. Reporting in academic writing serves a range of purposes: to present what has been done as well as what has not been done in a field; demonstrate the writer's familiarity with the field (Bavelas 1978); as tools to persuade (Gilbert 1977); to locate the writer's work in the larger context (Myers 1990) etc.. The writers carry out their own research based on reviewing prior works which are important background of new research. How to report and evaluate previous research properly and effectively becomes an indispensable part of academic writing.

In the teaching of academic writing in China, the importance and complexity of reporting are ignored by both language teachers and students. Even in some grammar books and language textbooks, reporting is regarded to be a very ordinary linguistic phenomenon which does not need to be defined. Some researchers (Bruce 1989; Granger 1993; Pickard 1995; Thompson & Ye 1999) found that non-native writers have more or less difficulties in reporting in their academic writing. Thomas & Hawes point out EAP learners commonly have serious difficulties with the range of choices involved in reporting such as choices of syntactic form, tense, voice and reporting verbs (1994, p.13). Previous studies indicate Chinese learners of English have some problems such as limited variety in reporting verbs, lack of critical thinking to prior research, misuse of reporting verbs, etc. Besides, I found it is difficult for them to report appropriately and effectively in their English theses, despite some of them being English majors or postgraduates who are proficient in English. Some Chinese learners of English are unaware of functions of reporting and make reference to previous works improperly. They even don't know how to use citations properly to increase credibility of their own claims. Other problems such as 'a strong tendency for non-native writers to overuse native quotation without interpretation' (Bruce 1989), 'inexpert paraphrasing which can cause confused or misleading indication of original meaning' (Groom 2000) often occur.

This study is primarily motivated by the difficulties and problems in reporting of Chinese learners of English. A lot of previous studies have investigated phenomenon of reporting in academic writing; however, these studies only focus on English-language writing, and few studies compared similarities and differences of reporting between Chinese-language and English-language. Moreover, although contrastive studies on reporting in English academic writing of native speakers and non-native speakers have been extensively carried out in recent years, few studies explored the reasons why Chinese learners of English behave differently in contrast with native speakers of English in reporting in English academic writing and few studies compared how Chinese writers and native-speakers-of-English report in their own languages in academic writing. The present study makes a comparison between reporting of Chinese writers and English writers in their thesis writing in their own languages, which may be a possible explanation to 'why Chinese learners of English behave differently compared with native-speakers-of-English in reporting in English academic writing'. The results may helpful to teaching of academic writing to Chinese learners of English.

1.2 Aim and objectives of this study

This study aims to investigate similarities and differences in reporting in master's theses of different disciplines between Chinese writers and English writers in their mother tongues. Besides, the disciplinary difference is also examined.

The objectives of this paper are to seek out the answers to the following research questions:

1. Are there any similarities and differences in reporting structure between Chinese writers and English writers in terms of integral-ness and prominence?
2. Are there any similarities and differences in reporting forms (short direct quotes, extensive direct quotes, summary, generalization, and list for reference) between Chinese writers and English writers?
3. How do Chinese writers and English writers differ in using reporting verbs in terms of categories of denotation and evaluation and usage of tense?

4. Do Chinese writers and English writers report in the same manner? Are there any differences in functions of reporting (background, support and critical)?

1.3 Organization of this paper

This paper consists of five sections. The introduction chapter introduces background and rationale of this study, and summarizes research questions and outline of this paper.

Chapter 2 is literature review which clarifies some important concepts in reporting and reviews some previous studies on reporting in academic writing.

Then methodology of this study is introduced in chapter 3, which describes data collection and data analysis in detail.

Afterwards, chapter 4 focuses on the results and discussion. In this chapter, answers to research questions raised in the Introduction section are provided though quantitative analysis.

Finally, conclusions are drawn to general findings of this study, limitations of this study stated and suggestions put forward for the future research.

Chapter 2 Literature review

2.1 Defining reporting

There are significant differences between reporting in academic writing and reported speech. Reporting in academic writing refers to academic writers making reference to previous works, including ideas (hypotheses, claims), research outcomes or findings, and research activities (research events/process) of previous research agents (individual or groups). What is reported can be referred to as a single study, as groups of studies, or as general level and trends of research (Benson et al. 2005, p.224). Reported speech refers to the account of the original speech event and it is used to report what an original speaker actually says. For example, *Labov (1961) conducted the first social dialect study on Martha's Vineyard.* This is reporting of a study of a sociologist, but it is not a reported speech.

Reported speech is studied in relation to both spoken discourse and written discourse while reporting in academic writing is studied in relation to written text. In terms of forms, reported speech generally includes direct speech and indirect speech while reporting in academic writing is in the forms of direct quotes, summary, paraphrase and generalization. Moreover, their functions are different. Sakita (2002) points out that the reported speech reflects the system and nature of communication, social dynamics and human cognition, and intrinsically manifests the dialogic nature of language (p.3). Reporting in academic writing serves a variety of purposes: to locate the writer's work in the larger context (Myers 1990); describing what has been done to indicate what has not been done in order to create a new research space (Swales 1990); to define a specific context of knowledge or problem (Hyland 1999); to give statements greater authority (Gilbert 1977).

Besides, focuses of the two items are different. Traditional reported speech can be studied from many perspectives. For instance, ethnographers study the use of reported speech as a social tool; linguists focus on the grammatical structure of reported speech (Sakita 2002, p.5) such as transformation from direct speech to indirect speech including tense and deixis shift, relationship between direct and indirect speech; while literary theorists turn to the significance of reported speech for narratorship (ibid). Different from traditional concept of reported speech, reporting in academic writing

can be investigated from these perspectives: frequency, categories and tense/voice usage of reporting verbs; integral and non-integral structure of reporting; prominence of reporting, forms of reporting; reporting signals, functions of reporting, characteristics of reporting in different genres or different disciplines; and contrastive analysis of different reporting practice of writers with different languages, etc.

Thomas & Hawes (1994) define reporting in academic writing as 'attribution of propositional content of a source outside the author of the article in the current situation, and the marking of this by presence of any of a number of signals of attribution' (p.129). According to Thomas & Hawes (1994), the article where the reporting occurs is the current/primary situation while the source article from which the reported comes is the previous/secondary situation (p.129). Reporting can be identified from a number of reporting signals: reporting verbs, reporting noun phrases, reporting adjectives, reporting adjuncts (including reporting adverbs, prepositional phrase, subordinative finite clause).

Citation is a kind of formal and explicit reference to the previous works in academic writing; however, reporting and citation are not identical although most citations are reporting. Consider the following examples:

a) Some scholars have found that learning strategies are crucial in second language learning.

   This reporting states a general trend in one area. It is not a citation but a summary as Jacoby (1987) proposes.

b) Kasper (1981) showed that native speaker norms in German and English differ. (source: the example is cited in Odlin, 1989)

c) Previous research has shown that native speaker norms in German and English differ (Kasper, 1981)

d) Native speaker norms in German and English differ (Kasper, 1981)

Examples b), c) and d) are both citation and reporting. Swales (1990) proposes that the examples '*The moon is probably made of cheese (Brie 1988).'* and *'According to Brie (1988), the moon is made of cheese.'* are non-reporting citations because Swales claims the reporting is introduced by a reporting verb, however, reporting verb is only

one signal of reporting. Reporting noun phrases, reporting adjectives, reporting adjuncts (including reporting adverbs, prepositional phrase, subordinative finite clause) are signals to identify reporting. Reporting in academic writing refers to making reference to previous research, including the ideas, findings and research activities of previous researchers. In the present study, the above examples b), c) and d) are all regarded as reporting.

2.2 Previous studies on reporting in academic writing

There has been growing interest in studies on practices of reporting since the publication of Swales' (1990) *Genre Analysis*. Many scholars have contributed a lot to study of reporting, for example, Thompson & Ye (1991); Thomas & Hawes (1994); Hyland (1999).

2.2.1 Structural studies of reporting

Swales (1981, 1986, 1990) is a pioneer in investigating reporting. He (1981) proposes a Four-move model (establishing the research field, reporting previous research, preparing for present research, and introducing present research) to analyze introduction of articles where reporting frequently occur but it is criticized by Crookes (1986) for not successfully to be applied to the literature in area of social science (cited in Swales 1986, p.45) and also criticized by some analysts for difficulty of separating Move 1 and Move 2 (cited in Swales 1990, p.140). Later, Swales (1990) revised his Four-move model into CARS (create a research space) pattern (three moves: establish a territory, establishing a niche, and occupying the niche). Swales (1990) also makes a distinction between integral and non-integral citation: the former refers to citation in which the names of researchers appear in the citing part as a grammatical element while the later refers to citation in which the names of researchers occur outside the citing sentence either in the parenthesis or other device (p.148). Swales' distinction between integral and non-integral citations becomes a solid foundation for studying reporting; however, he does not clarify the function of each citation type and thus provide no insights that help academic writers or student

writers understand which type to use in which context (Thompson & Tribble 2001, p.92). Besides, his CARS pattern does not account for all the irregularities and variety found in authentic journal writing (Pickard 1995, p.93). Most importantly, Swales' integral and non-integral distinction is not complete enough, which cannot be applied to all the reporting phenomena. For example, a general statement which states the general trend of previous research as a whole or a reference to consensus knowledge cannot be classified by integral or non-integral citation. E.g. *An increasing number of researchers in all areas of psychology have recognized the critical role of culture in shaping psychological processes.* Besides, other reference types such as repeated name of previously mentioned researcher, a reporting sentence without citation and mention of a common recognized theory, model or law cannot be analyzed by integral or non-integral citation either.

To understand how previous research is reported, integral and non-integral citations should be related with prominence (Weissberg & Buker 1990). According to citation focus, Weissberg and Buker (1990) propose information prominent citation, author prominent citation, weak-author prominent citation and general statement. If information is given primary importance and author's name and year of publication (in some cases, citation number referring to the numbered list of references at the end of the paper is used) is attached in the brackets at the end of the sentence, the citation is called information prominent citation. If the author's name is given more focus as the subject of the sentence followed by the date or citation number in parentheses, it is author's prominent. Weak author prominence refers to the citation in which un-named group serves as the subject or agent in the sentence, or follows the prepositional phrases such as *according to*. General statement is used to describe the level of research activity in an area (Weissberg & Buker 1990). However, this classification is not perfectly complete. Benson, Gollin & Trappes-Lomax (2005) added 'author + information' reference type in which author's name as part of a possessive noun phrase. Swales' integral-ness (1990) and Weissberg and Buker's prominence (1990) are combined to analyze reporting in the present study.

Integral structure of reporting is also studied by Pickard. Pickard (1995) explores how "expert" writers use citations by using a small corpus of 11 applied linguistics articles.

She conducted a concordance search to investigate the use of integral and non-integral citation by "expert" writers and also identified four grammatical forms of integral citation based on Swales' (1990) division of integral and non-integral citation. Although Pickard's research provides detailed categories for further studies on integral-ness, there are some limitations. These different categories are classified in terms of syntactic perspective but their functional aspects are not mentioned at all. Little suggestion is given for academic writers to choose which form in which context. Furthermore, the sample was from applied linguistics. She does not clarify whether the findings of citation practice of "expert" writers in applied linguistics can be generalized to ones in other disciplines as the study of Hyland (1999) indicates that there are differences across disciplines in how writers use citations in academic writing.

2.2.2 Studies on reporting forms

How previous works are referred to is another aspect of reporting that arouses interest of researchers (Dubois 1988; Thompson 1996; Hyland 1999). Swales (1986) created *short* and *extensive* citations according to the length of citation: short citations refer to the sentential citations and extensive citations encompass more than one sentence (p.50). Dubois (1988) examined how scientists use cited works in biomedical journal articles by comparing source articles and citing articles. Dubois introduces four forms of reporting: direct quotation, paraphrase, summary and generalization. According to Dubois, 'paraphrase is restatement of an idea in different words but the same length, summary is an abbreviated statement of a result or fact from a single source article, and generalization is a statement of similarity from the work of two or more source articles' (p.183). Dubois found most instances of reporting are in the form of summary and generalization, which is consistent with Hyland's (1999) findings. However, different categories of forms are employed in Hyland's study: short direct quotes (three or more words), extensive quotes as blocks, summary from a single source, and generalization from two or more sources. In the present study, short direct quotes, extensive direct quotes, summary, generalization and list for reference are examined. Paraphrase is not discussed in this study because only citing articles are examined. Whether a reporting sentence is paraphrase or not can only be identified by

comparing the source articles and citing articles; however, in the present study only citing articles (80 theses) are examined. List for reference is added to forms of reporting because Dubois' (1988) and Hyland's (1999) classifications are not watertight. List for reference is another reference type in many research articles. List for reference refers to mention of some authors in brackets as un-named group or mention of some source articles as examples, comparison, detailed references (e.g./cf./see).

2.2.3 Studies on reporting verbs

Alongside Swales' early work on reporting, intensive studies have been carried out to investigate reporting verbs as the most prominent signal of reporting (Malcolm 1987; Shaw 1992; Thompson & Ye 1991; Thomas & Hawes 1994). Most researchers focus on the categories of reporting verbs, tense and voice of reporting verbs with sentence function.

A lot of studies have analyzed tense usage of reporting verbs (Oster 1981; Een 1982; Hanania & Akhtar 1985; Malcolm 1987; Swales 1990; Shaw 1992). These studies have examined the use of present, past tense and present perfect. Lackstrom, Selinker and Trimble (1972) conclude that present tense indicates a general claim, past tense claims lack of generality and present perfect tense gives a good generalization about past events (cited in Shaw 1992, p.303). Malcolm (1987) holds the similar ideas and she analyzed tense choice in 20 scientific articles from context-independent temporal meanings and context-dependent rhetorical uses. Malcolm (1987) found that generalizations tend to occur in present tense, reference to specific experiments in the past tense and reference to areas of inquiry in the present perfect tense (p.36). Malcolm's (1987) study throws light on the implications for the teaching of EST; however, this study only analyzed isolated clauses and the data is relatively small (p.41). Moreover, Een （1982）found that Malcolm's claim that reference to specific experiments occurs in the past tense does not account for all the data. In Contrast with Malcolm's (1987) findings, Oster (1981) generalizes her hypotheses:

> The present tense is used primarily to refer to quantitative results of past literature that are supportive of or non-relevant to the work in the current article.

The past tense primarily claims non-generality about past literature and secondarily refers to quantitative results of past literature that are non-supportive of the work in the current article.

The present perfect tense is primarily used to indicate continued discussion of some of the information in the sentence in which the present perfect tense occurs, and secondarily used to claims generality about past literature.

(Oster 1981, p.77)

Oster's (1981) analysis of tense usage is associated with the nature of the claims being made about the previous literature. Although Oster's sample is very small (only two articles), her hypothesis of present perfect is explained from discoursal perspective rather than semantic/sentential perspective (cited in Swales 1990, p.152). Swales (1981, 1990) found it difficult to decide which part count as 'continued discussion' and he proposes that past tense occurs frequently in the integral reporting sentence with the name of researcher as subject, present perfect in the non-integral reporting sentence, and present tense in the non-integral citations. Swales (1990) also suggests that the choice of tense may indicate the writer's stance towards the cited work (p.154). Swales and Feak (1994, 2004) argues there are three major patterns:

Pattern I --- reference to single studies---past

Pattern II --- reference to area of inquiry---present

Pattern III --- reference to state of current knowledge---present perfect

They also point out moves from past to present perfect then to present indicate that the research reported is increasingly close to the writer's own opinion or research.

Weissberg and Buker (1990) analysed tense usage of reporting verbs in terms of prominence:

1) The past tense is used in the findings of individual studies closely related to you own;
2) The present tense is used in the information prominent citations when the cited information is generally accepted as scientific fact.
3) The present perfect tense is used in weak author prominent citations and general statements which describe the level of research activity in an area.

(Weissberg & Buker, 1990, p. 51-52)

Furthermore, Weissberg and Buker (1990) note correlation of attitude and tense in the reported findings:

1) Past tense is used in the findings which you believe are restricted to the specific study you are citing but not be acceptable as true in all cases;
2) Present tense is used in the findings which you believe are fact;
3) Tentative verbs and a modal auxiliary with the complement verb are used in the findings you are citing were considered by the original author as tentative, or were only suggestions or proposal rather than findings.

(Weissberg & Buker, 1990, p.55-56)

Compared to Swales and Feak's (1994, 2004) patterns, Weissberg and Buker's patterns are more comprehensive and elaborate.

Later, Shaw (1992) examines how tense of reporting verbs is used in Ph.D theses and explores the reasons of correlation of tense and sentence function. Shaw points out topicalization and topic change should be considered when analyzing the tense of reporting verbs. Findings of Swales & Feak (1994, 2004) and Shaw (1992) are similar.

In summary, past tense is used when referring to a specific study or experiment which may be close to your current study, and the findings of the study or experiment are limited to the cited study. Present tense is used in the findings which are believed as fact or supportive of the current study. Present perfect is often used in generalization of research activity in an area or used to indicate continued discussion in the current study.

In regard to semantic categories of reporting verbs, the most notable early research is Thompson and Ye's (1991) evaluation in reporting verbs used in academic papers. Thompson and Ye classify reporting verbs in terms of denotation and evaluation. In analysis of denotation, they propose three categories: textual, mental and research verbs which are under the heading 'author acts' (following Thompson & Ye, we use 'writer' to refer to the person who is reporting and 'author' to refer to the person who

is being reported):

> Textual: verbs referring to processes in which verbal expression is an obligatory component; for example, state, write, point out, term, deny, etc.
>
> Mental: verbs referring primarily to mental process; for example, believe, think, focus on, consider, etc.
>
> Research: verbs referring primarily to the mental or physical processes that are part of research work (and to the author's descriptions of those processes); for example, measure, obtain, find, calculate, etc.

(Thompson & Ye 1991, p.369-370)

In analyzing the evaluative nature of reporting verbs, they consider three factors: author's stance, writer's stance and writer's interpretation. We focus on author's stance and writer's stance in the present study. Three options are identified by Thompson and Ye (1991):

> Factive: the writer portrays the author as presenting true information or a correct opinion; for example, demonstrate, points out, identify, prove, improve, notice, etc.
>
> Counter-factive: the writer portrays the author as presenting false information or an incorrect opinion; for example, betray, confuse, disregard, ignore, use, etc.
>
> Non-factive: the writer gives no clear signal as to her attitude towards the author's information/opinion; for example, believe, claim, examine, propose, generalize, utilize, etc.

Although this is a useful study, it is not watertight. Some reporting verbs can be classified into two categories because the distinction between each category is not so easily distinguishable. Following Thompson and Ye's (1991) study, Thomas and Hawes (1994) analyze reporting verbs in medical journals by looking at a small corpus of 11 research articles. They focus on the semantic categories of reporting verbs and identify function of reports with each category. They categorize denotation of reporting verbs in terms of experimental/real-world activities, cognition activities and discourse activities. This classification is similar to Thompson and Ye's classification of research, mental and textual verbs. This study provides useful

insights that suggest there is a correlation between choice of verb type and the function of the report in which the verb occurs (p.147); nevertheless, their corpus is relatively small and is restricted to one discipline and the modified categories of denotation are basically same as Thompson and Ye's (1991). Based on the work done by Thompson and Ye's (1991) and Thomas and Hawes (1994), Hyland (1999) investigated a corpus of 80 research articles of eight disciplines. Hyland regards Thompson and Ye's (1991) classification to be an over-complex system and there is no need to distinguish evaluation of reporting verbs in ten sub-categories (p.350). Therefore, Hyland simplified this system by categorizing evaluation in terms of factive, non-factive and counter-facitve.

To sum up, reporting verbs can be analyzed under headings of denotation and evaluation proposed by Thompson and Ye (1991). Denotation and evalution are all classified from perspectives of the author and the writer. The classification of Thompson and Ye (1991) is employed in the present study.

In China, some scholars have become interested in studying reporting verbs in recent years. He and Zhou (2001) examined semantic categories of reporting verbs in 108 academic articles from nine disciplines while Hu and Jiang (2007) conducted a contrastive research on reporting verbs in 13 English M.A. theses of Chinese learners of English and native speaker of English. Both of the two studies are a corpus-based quantitative studies and employed Thompson and Ye's (1991) categories of reporting verbs, but the former discussed reporting verbs from perspective of semantics and pragmatics while the latter mainly investigated different preferences of Chinese learners of English and native speaker of English in using reporting verbs. There are some limitations in two studies. He and Zhou (2001) use Gosden's (1993) classification of subjects to analyze reporting verbs; however, He and Zhou do not clarify the correlation of choosing reporting verbs and their subjects from perspectives of semantics and pragmatics. Although Hu and Jiang's (2007) contrastive study is a preliminary attempt, the corpus is too small to draw strong conclusions.

2.2.4 Studies on reporting in different disciplines

Some studies focus on the disciplinary differences in reporting. Hyland (1999) examined reporting practice in a corpus of 80 research articles of eight disciplines and identified across-discipline variation. Another similar study is Thompson's (2000). Both of the two studies use large-sized corpora to analyze reporting in different genres of academic writing. Thompson (2000) investigated reporting in fourteen PhD theses of native-speakers-of-English by using a tagging system and concordancing techniques. By examining doctoral theses from Department of Agricultural and Food Economics and Department of Agricultural Botany, Thompson (2000) devised a set of functional categories based on Swale's (1990) distinction of integral and non-integral citations and Thompson and Ye's (1991) classification of denotation of reporting verbs. The results of Thompson's study reveal that there are distinct differences in the use of citations in two sub-disciplines and indicate the writers in two sub-disciplines construct texts in markedly different ways (p.100). Hyland (1999) found citation practices are different in 'hard' and 'soft' disciplines proposed by Becher (1989) and Kolb (1981). Hyland believes the distinction between hard and soft disciplines provides a useful basis for analyzing disciplinary differences; however, Hyland also points out this distinction is inadequate and only acceptable at a general level of analysis. Hyland's (1999) investigation shows disciplinary differences in the extent to which writers refer to the cited works in presenting arguments and in how they choose to represent such work (p.346). Later, Thompson and Tribble (2001) made an elaborate comparison of Hyland's (1999) and Thompson's (2000) studies and put forward some suggestions for EAP teachers and students in investigating of citation practice in genres.

2.2.5 Studies on function of reporting

The functions of reporting also attract researchers' attention. Weissberg and Buker (1990) propose reporting has three functions: first, giving readers background information about your study; second, showing readers your familiarity with the area; third, establishing your study as one link in a chain of research that is developing and enlarging knowledge in your field (p.41). The second function is similar to Bavelas' (1978) claim that references to prior works demonstrate that you are familiar with the

important work in the field (p.160). Bavelas also mentions that citations 'supply evidence that you qualify as a member of the chosen scholarly community' (cited in Swales & Feak 1994, p252). However, showing familiarity in the field is not the primary function of making reference to previous works in research paper and master's or PhD's theses. Swales (1990) points out reporting of previous research is to create a research space by describing what has been done and what has not been done. Gilbert (1977) considers references to previous works as 'tools of persuasion' and 'demonstration of validity and significance of the work reported in the scientific papers' (p.115-116). Later, Bloch and Chi (1995) propose four functions of reporting: background, support, faulty path and return path. The concept of faulty path and return path will be explained below. Gilbert's (1977) views are consistent to function of 'support' in Bloch and Chi (1995) and Swales' (1990) views are similar to the function of 'background' and 'faulty path and return path'. The classification of function of reporting proposed by Bloch and Chi (1995) is adopted in the present study.

1) Background: references are not directly related to the argument the writer is making, including citations referring to methodology, definitions, explanations and historical references presented uncritically.
2) Support: citations directly related to supporting the argument the writer is making or supporting a point the writer is making.
3) Faulty path: citations the writer disagree with either partly or completely.
4) Return path: citations support points of the writer's disagreement.

(Bloch & Chi 1995, p.242)

As Bloch and Chi point out, it is difficult to distinguish exactly whether the citation is used as a faulty path or a return path, and the two categories are both concerned with the critical analysis of prior texts, they propose the category 'critical' instead of a faulty path and a return path. Therefore, three functions of 'background, support and critical' are employed in this study.

2.3 Summary

Although previous studies contribute a lot to the study of reporting in academic writing, the majority of these studies only focus on English-language. Bloch and Chi (1995) compared citations between Chinese-language and English-language, but they only focused on two aspects: the comparison of the time period of the source texts, and how Chinese writers and English-language writers differ in using citations. Few studies made a thorough comparison between how Chinese-language writers and English-language writers report from perspectives of reporting structure, reporting forms, reporting signals, and reporting functions, etc. The present study investigates similarities and differences between how Chinese writers and native speaker of English writers make reference to previous works in their academic writing of their own languages in terms of integral-ness and prominence, reporting forms, semantic categories and tense usage of reporting verbs, and reporting functions.

The previous project I have done in the course *Topics in Grammar and Discourse* investigated how Chinese writers and native speakers of English differ in reporting in English thesis writing. The study found that differences indeed exist between CLE and NS, for example, CLE do not use reporting verbs or integral citations as much as NS. The focus of previous project is on the English theses of two different groups while the present study focuses on theses of two languages of two different groups. Based on the previous project, the present study explores how these two different groups differ in reporting in their native language thesis writing, which may be a possible explanation to 'why Chinese learners of English behave differently from native speakers of English in reporting in English academic writing'.

Chapter 3 Methodology

3.1 Data collecting

The previous studies (e.g. Bloch 1990; Hyland 1999; Thompson 2000) indicate there are differences across disciplines in how writers use reporting in academic writing. So the sample in this study is collected from different disciplines of two areas: social sciences (including psychology, philosophy, economics, and history) and physical sciences (including biology, physics, electronic engineering, and chemistry).

English theses were collected from internet resources of some universities such as University of Edinburgh, MIT, University of Toronto, etc. Chinese theses were collected from internet resources of key universities in mainland China, such as Southeast University, Tsinghua University, Shanghai Jiaotong University, etc.(Other detail about data is listed appendix Ⅰ). 40 English theses written by native speakers of English and 40 Chinese theses written by native Chinese writers are chosen, among which 5 articles are selected from each discipline. These theses were published in recent years. English corpus and Chinese corpus are established respectively. Table 3.1 shows the details of the two corpora.

Table 3.1 Details of Chinese corpus and English corpus

Areas	Disciplines	English theses	Chinese theses
Social sciences	Psychology	5	5
	Philosophy	5	5
	Economics	5	5
	History	5	5
Physical sciences	Biology	5	5
	Physics	5	5
	Electronic engineering	5	5
	Chemistry	5	5
Total	80	40	40

3.2 Data analysis

Reporting usually occurs in the sections of Introduction and Literature review; therefore only Introduction and Literature review are examined in the present study. These two sections of each thesis are read line by line and are tagged with a tagging system. Wordsmith Tools are used to examine these tagged items for further analysis because WST are powerful in searching and concordance, which can reduce the time-consuming manual counting work.

The word counts for the two corpora and two areas are shown in the following tables. As mentioned before, only sections of Introduction and Literature review are counted.

Table 3.2 Word counts in Chinese corpus and English corpus

Corpus	Word counts
Chinese corpus	228420
English corpus	118020

Table 3.3 Word counts in social sciences and physical sciences

Areas	Word counts
Social sciences	152190
Physical sciences	194250

It should be pointed out the word counts are approximate numbers rather than exact numbers because it is difficult to count words of some theses due to their file types.

Data analysis for Question 1 integral-ness and prominence

In addition to integral-ness (Swales 1990), prominence (Weissberg & Buker 1990), and 'author + information' (Benson, Gollin and Trappes-Lomax 2005), three reference types are added in the present study: a specific reporting without citation; repeat name of cited researcher mentioned above without citation and a common recognized theory, model, law etc. without citation. These are explained later when they are tagged.

Jacoby (1987) notes summary is a kind of reference in which no particular previous researcher is named but clear reference to the state of previous research as a whole or the state of consensus knowledge can be identified (p.55). This category of reference is similar to 'general statement' proposed by Weissberg and Buker (1990). These general statements are often written without citations, but sometimes the general statements occur with mention of some researchers as examples, e.g. *In the wider perspective of research into second language acquisition (e.g. Faerch and Kasper; Rost 1990), listening is regarded as a powerful source of input…* (cited in Gollin & Trappes-Lomax 2005, p.226). Therefore, there are two subcategories under 'general statement': one is general reporting tagged with <re_gs> and the other is non-integral citation tagged with <non_gs>.

Table 3.4 The tagging system for question 1

	Prominence	Type of citation	Illustration	Tagging
1	Information prominent	Non-integral	Information + (reference or reference number)	<non_ip>
2	Author prominent	Integral	1) Author's name as subject 2) Author's name as agent (by/in/from) 3) Prepositional phrases such as according to, in accordance with, as…points out	<in_ap1> <in_ap2> <in_ap3>
3	Author and information	Integral	Name of author as part of possessive noun phrase	<in_ai>
4	Weak author prominent	Non-integral	Reference to un-named research group	<non_wap>
5	General statements	Non-integral citation	Reference to a general description of the state of previous research with mention of some researchers as examples	<non_gs>
		Not citation	Reference to a general description of the state of previous research without citation	<re_gs>
6	reporting	Not citation	Specific reference without citation	<re>
7	Others	Not citation	1) Reference to repeat name of cited researcher mentioned previously 2) A common recognized theory, model, law etc.	<re_rep> <re_com>

As shown in Table 3.4, categories according to integral-ness and prominence are tagged with different symbols:

1. Information prominent:

   Tag: <non_ip>

In information prominent citations, the information is given primary emphasis and the name of author and the date of publication (sometimes also with page number) are put in the brackets at the end of the sentence. Another type is that the reference number is used instead of author's name and publication year. The number refers to the list of references in the reference section at the end of the paper (Weissberg & Buker 1990). Here are some examples from the sample:

> Some individuals exhibit weaker language lateralization which means their language skills are spared after a unilateral lesion (Knecht et al. 2002).
>
> Due to the importance of this chromophore, innumerable experimental and theoretical studies have been carried out on ethylene and other olefins[1-10].

2. Author prominent:

As to author prominent, the author's name is given primary importance. There are some sub-types under this category:

1) Author's name as subject of the sentence

   Tag: <in_ap1>

   E.g. Frisch (1996) proposes that the co-occurrences of certain feature combinations results in redundant information, such that information about a feature configuration varies in importance depending on the configuration of the other features.

2) Author's name as agent (by/in/from)

   Tag: <in_ap2>

   E.g. Subsequent work by Richardson et al. (2003) investigated this phenomenon and found subjects were more likely to remember an emotionally arousing word such as *murder* or *scream* than a more neutral one, such as *carpet* or *block*.

3) Prepositional phrases such as according to, in accordance with, as…points out, following…

Tag: <in_ap3>

E.g. According to Nielsen (1999a), navigation should be placed on the left side as it is common convention.

3. Author and information

Tag: <in_ai>

E.g. This does not disparage Beeman's (1993) coarse and fine coding theory in any way.

4. Weak author prominent

Tag: <non_wap>

E.g. However, a large number of researchers (Hill and Rothaermel 2003, Uttweback 1994, Foster 1986, Christensen 19997, Henderson and Clark 1990, Cooper and Schendal 1976), studying a wide range of industries observed at different periods in history….

5. General statements

1) Non-integral citation

Tag: <non_gs>

E.g. Modern theories of language production are by and large all based around the same representational level, starting at the high level semantics of a word, and extending downwards through its syntactic (or lexical), phonological and phonetic representations towards final articulatory output (e.g. Dell, 1986; Levelt, 1989; Levelt, Roelofs & Meyer, 1999).

2) General reporting

Tag: <re_gs>

E.g. An increasing number of researchers in all areas of psychology have recognized the critical role of culture in shaping psychological processes.

6. Specific reporting

Tag: <re>

E.g. In 1959 Deese discovered that when presented with a list of words to recall later, subjects could be induced to falsely remember a non-presented critical word.

7. Others

1) Repeat name of cited researcher mentioned above

Tag: <re_rep>

E.g. This current research is a replication and extension of Ito's (2001) who examined hemispheric asymmetry in the induction of false memories. Ito used a false recognition paradigm and a standard list learning paradigm in order to investigate hemispheric asymmetry in verbal memory recall.

2) A common recognized theory, model, law etc.

Tag: <re_com>

E.g. … "paternalistic tort law" is determined to be an oxymoron.

Data analysis for question 2: reporting forms

In this study, four categories based on Dubois (1988) and Hyland (1999) are examined: short direct quotes, extensive direct quotes, summary, generalization and list for reference.

1. Short direct quotes

Tag: <sdq>

Short direct quotes refer to sentential quotes, usually around three words.

E.g. Leading up to self-harm, people often report feeling rejection, depression, restlessness, tension, "…then finally numbness, emptiness, and self-absorption…" (Zila & Kiselica, 2001, 79).

2. Extensive direct quotes

Tag: <edq>

E.g. They note: "By definition, the large semantic fields activated by RH coarse semantic coding are only weakly activated, and weak facilitation of the target word might be difficult to detect - just as weak RH semantic processing of single words is difficult to observe" (Beeman et al. 1994, p.29).

3. Summary from a single source

Tag: <sum>

Summary in the present study includes paraphrase. Paraphrase is not examined separately in this study because paraphrase cannot be identified unless both the source articles and citing articles are compared. Only citing articles are examined in this study.

E.g.a) Individualists and collectivists differ in the kinds of sociability they prefer, the meaning of social interactions, and their beliefs about important groups (Osyerman et al. 2002).

b) Jackson and Smith (1999) found that Americans high in COL showed higher levels of in-group pride compared to those low in COL.

4. Generalization from several sources

Tag: <gen>

E.g. a) This list learning paradigm…has subsequently been used in other experiments to explore, amongst other things, converging associative networks (Watson, Balota & Roediger 2003); memory illusions, (Roediger 1996); long lived semantic priming (McKone and Murphy, 2000); and even memory distortion in individuals who claim to have been abducted by aliens (Clancy et al. 2002).

b) Immunosuppression is another mechanism parasites can employ to escape the harmful effects of the immune response to foreign in vaders (Cohen 1976; Barriga 1984; Piessens et al. 1980; and Jayavardena & Wakesman 1977).

5. List for reference

Tag: <list>

List for reference refers to mention of some authors in brackets as un-named research group or mention of some specific source articles as examples, comparison, detailed references (e.g./cf./see/see e.g.).

E.g. To support suggestions, convincing evidence of the lexical bias has been found through a number of experimental studies (e.g., Baars, Motley, & MacKay 1975; Dell, 1986, 1990; Hartsuiker, Corley & Martensen 2005)

Date analysis for question 3: reporting verbs

Evaluative potential of reporting verbs are examined from two aspects: categories (author's stance and writer's stance) and tense.

1. Categories of reporting verbs

Following Thompson and Ye (1991), classification of reporting verbs is list in the following tables.

Table 3.5 Tagging for categories of denotation

	Categories	Examples	Tagging
Author acts	Textual	propose, point out	<t>
	Mental	believe, think	<m>
	Research	find, discover	<r>
Writer acts	Comparing	correspond to	<c>
	Theorizing	support, account for	<th>

Table 3.6 Tagging for categories of evaluation

	Categories	Examples	Tagging
Author's stance	Positive	point out, emphasize.	<pos>
	Negative	oppose, challenge.	<neg>
	Neutral	examine, focus on.	<neu>
Writer's stance	Factive	identify, notice.	<fac>
	Counter-factive	ignore, misuse.	<cfac>
	Non-factive	claim, believe.	<nfac>

2. Tense of reporting verbs

Present Tag: <pre>

Past Tag: <past>

Present perfect Tag: <pp>

Data analysis for question 4: functions of reporting

The function of reporting is based on Bloch & Chi's (1995) classification of function of citation: background, support and critical.

Background Tag: <bg>

Support Tag: <sup>

Critical Tag: <cri>

Here are tagged examples taken from data:

> **<re> <sum> <bg>** In 1959 Deese discovered **<r> <fac> <past>** that when presented with a list of words to recall later, subjects could be induced to falsely remember a non-presented critical word.

<in_ap1> **<sum>** **<bg>** Clancy et al. (2002) argue **<t>** **<nfac>** **<pre>** that the DRM paradigm acts as a type of source monitoring error...

…**<non_ip>** **<sum>** **<cri>** However, recent developments in imaging techniques of the brain and new research methods are resulting in mounting evidence for sex differences not only at the structural level, but right down to the cellular level of the brain (Cahill, 2005)…

Chapter 4 Results & discussion

4.1 Integral-ness and prominence

I shall begin by comparing the Chinese corpus and the English corpus in terms of integral-ness and prominence.

Table 4.1 Frequency of integral-ness and prominence in Chinese corpus and English corpus

Prominence		Type of citation	Corpus of Chinese theses		Corpus of English theses	
			Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Information prominent		Non-integral	555	24.3	757	64.1
Author prominent		Integral	218	9.5	167	14.2
Author and information		Integral	43	1.9	30	2.5
Weak author prominent		Non-integral	14	0.6	22	1.9
General statements		Non-integral citation	6	0.3	11	0.9
		Not citation	30	1.3	10	0.8
Reporting		Not citation	16	0.7	24	2.0
Others	Repeated name of previously mentioned researcher	Not citation	12	0.5	59	5.0
	A common recognized theory, model, law	Not citation	2	0.09	6	0.5
Total			896	39.2	1086	92.0

Both raw counts and normed counts are included in Table 4.1. Because the word counts of the two corpora are different and raw counts only show the numbers of occurrences of each category, the frequency cannot be compared by raw counts. The numbers of occurrences in two corpora should be normalized before comparing. Normed count is used to make comparison because normed count is actual frequency which stands for occurrences of each category per 10000 words.

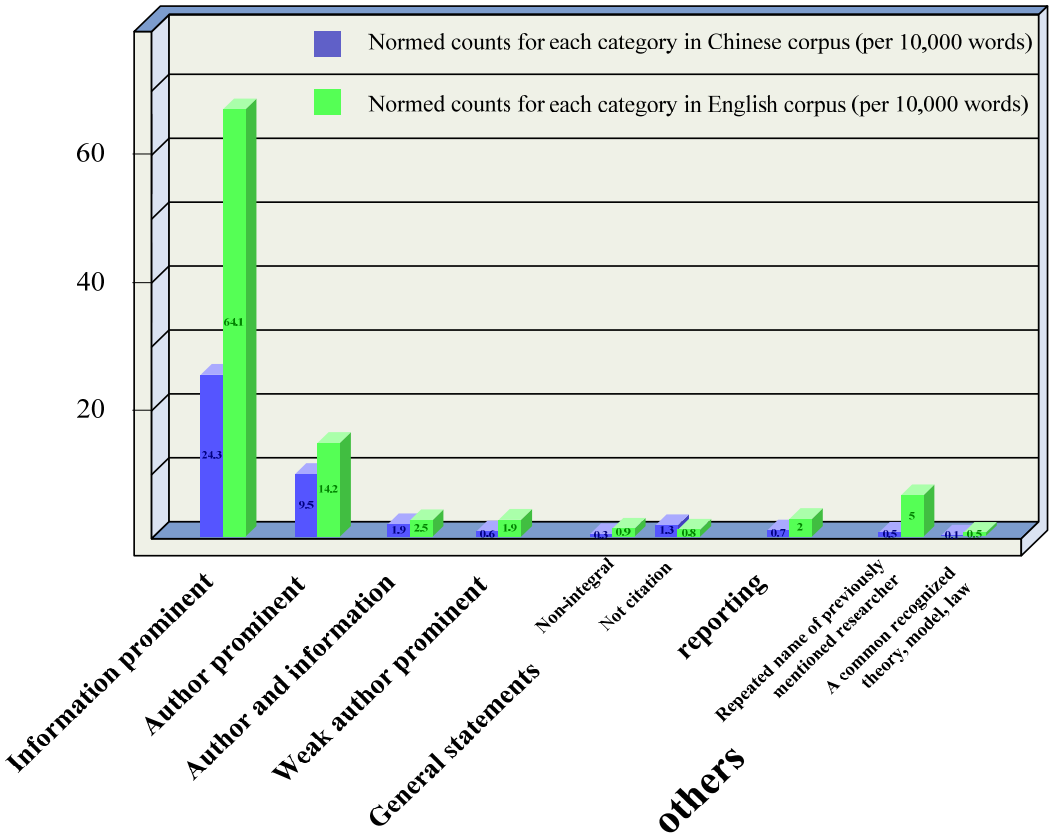

Figure 4.1 shows frequency of integral-ness and prominence in Chinese corpus and English corpus

As shown in Figure 4.1, there are more normed counts of nearly all reference types (except non-citation general statements) in English corpus than in Chinese corpus. The difference of information prominent citations between two corpora is most evident. Differences of other reference types like author and information prominent citations, weak author prominent citations, general statements, specific reporting and mention of a common recognized theory between two corpora are small. It can be also concluded that writers of English corpus use much more reporting statements than writers of Chinese corpus.

Table 4.2 Frequency of integral and non-integral citations in Chinese corpus and English corpus

Corpora	Approx. number of words in sample	Raw counts of integral citations	Normed counts for integral citations (per 10000 words)	Raw counts of non-integral citations	Normed counts for non-integral citations (per 10000 words)
Chinese corpus	228420	261	11.4	575	25.2
English corpus	118020	197	16.7	790	66.9
Total	346440	458	13.2	1365	39.4

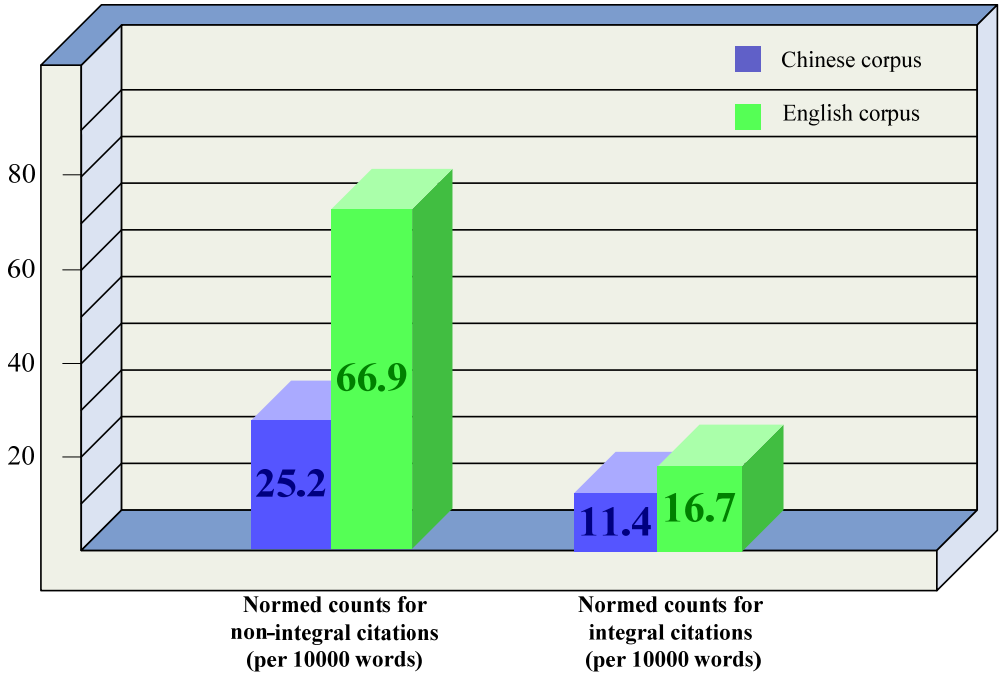

Figure 4.2 shows frequency of integral and non-integral citations in Chinese and English corpus

Figure 4.2 shows that Chinese writers use 11.4 integral citations per 10000 words in Chinese theses while native speakers of English writers use 16.7 integral citations per 10000 words. It is obvious that there are fewer integral citations in Chinese corpus than in English corpus. In regard to non-integral citations, there are about 41.7 per 10000 more words in English corpus than in Chinese corpus. The option of integral

and non-integral citation is related to the focus of reporting. It is associated with prominence. Integral citations are used in author prominent citations and author & information citations where focus is put on previous researchers. Non-integral citations are used when information is emphasized in information prominent citations, in weak author prominent citations and sometimes in general statements. That is to say, there are more author prominent citations and author & information citations in English corpus than in Chinese corpus. So is the same case with information prominent citations, weak author prominent citations and general statements. Generally speaking, there are more reporting statements used in English corpus than in Chinese corpus.

Next, I shall examine the proportion of integral and non-integral citations to all citations in sample.

Table 4.3 Proportion of integral and non-integral citations to all citations in both Chinese data and English data

Integral-ness	Raw counts	Percentage
Integral citations	458	25.1%
Non-integral citations	1365	74.9%
Total	1823	100%

Seen from Table 4.3, integral citations share a percentage of 25.1% while non-integral citations share 74.9%. There are much more non-integral citations than integral citations in total citations. Both writers in Chinese corpus and English corpus use non-integral citations more frequently.

In the end, I shall investigate disciplinary differences and proportion according to language and area.

Table 4.4 Frequency of integral-ness and prominence
in social science and physical science

Prominence		Type of citation	Data of Social sciences		Data of Physical sciences	
			Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Information prominent		Non-integral	502	33.0	810	41.7
Author prominent		Integral	277	18.2	108	5.6
Author and information		Integral	59	3.9	14	0.7
Weak author prominent		Non-integral	34	2.2	2	0.1
General statements		Non-integral	8	0.5	9	0.5
		Not citation	31	2.0	9	0.5
reporting		Not citation	109	7.2	31	1.6
others	Repeated name of previously mentioned researcher	Not citation	66	4.3	5	0.3
	A common recognized theory, model, law	Not citation	5	0.3	3	0.2
Total			1091	71.7	991	51.0

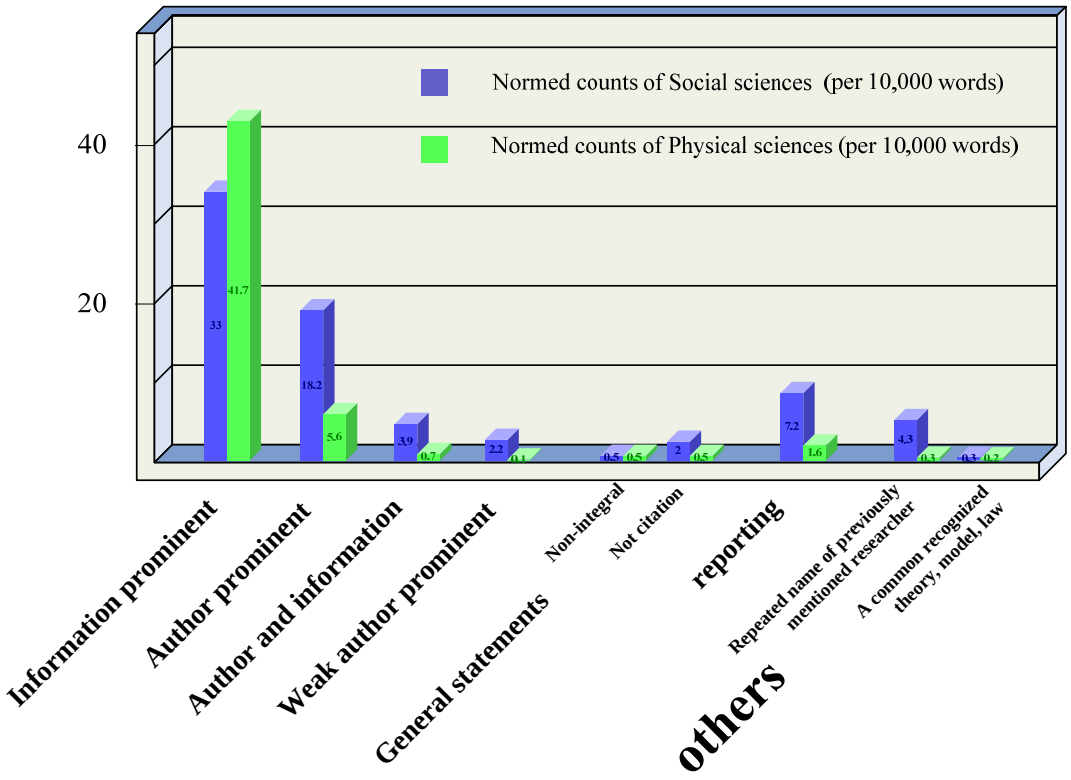

Figure 4.3 shows frequency of integral-ness and prominence in social science and physical science

As can be seen from Figure 4.3, there are more integral citations in theses of social sciences than in those of physical sciences while there are more non-integral citations in physical sciences than in social sciences. In non-integral citations, the most obvious difference is information prominent citations between social sciences and physical sciences. Writers of physical sciences use about 41.7 information prominent citations per 10000 words while writers of social sciences use 33.0 per 10000 words. A possible explanation for this is that in some disciplines such as physics, electronic engineering, writers make reference to some theories and concepts as procedures or tools to carry out their research, so they focus on information rather than researchers who invent or modify these methods or theories. However, in social sciences, disciplines such as philosophy and psychology, writers often give greater prominence to cited authors because themes of some theses of philosophy are analysis or disagreement of some theories of cited authors and cited authors are more emphasized.

Therefore, there are more integral citations in social sciences than in physical sciences. Writers of social sciences and those of physical sciences use almost the same amount of general statements. More repeated names of previously mentioned researcher occur in the theses of social sciences than in those of physical sciences. The possible reason for that is cited authors are given greater emphasis in social sciences as mentioned above.

Table 4.5 Frequency of integral-ness and prominence in different disciplines in social sciences and physical sciences

Areas	Disciplines	Integral		Non-integral	
		Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Social sciences	Psychology	243	51.3	294	62.0
	Philosophy	26	11.2	47	20.2
	Economics	48	10.1	121	25.5
	History	19	5.6	80	23.5
Physical sciences	Biology	62	8.4	455	61.3
	Physics	10	3.3	112	36.8
	Electronic engineering	1	0.3	66	19.7
	Chemistry	49	8.7	188	33.5

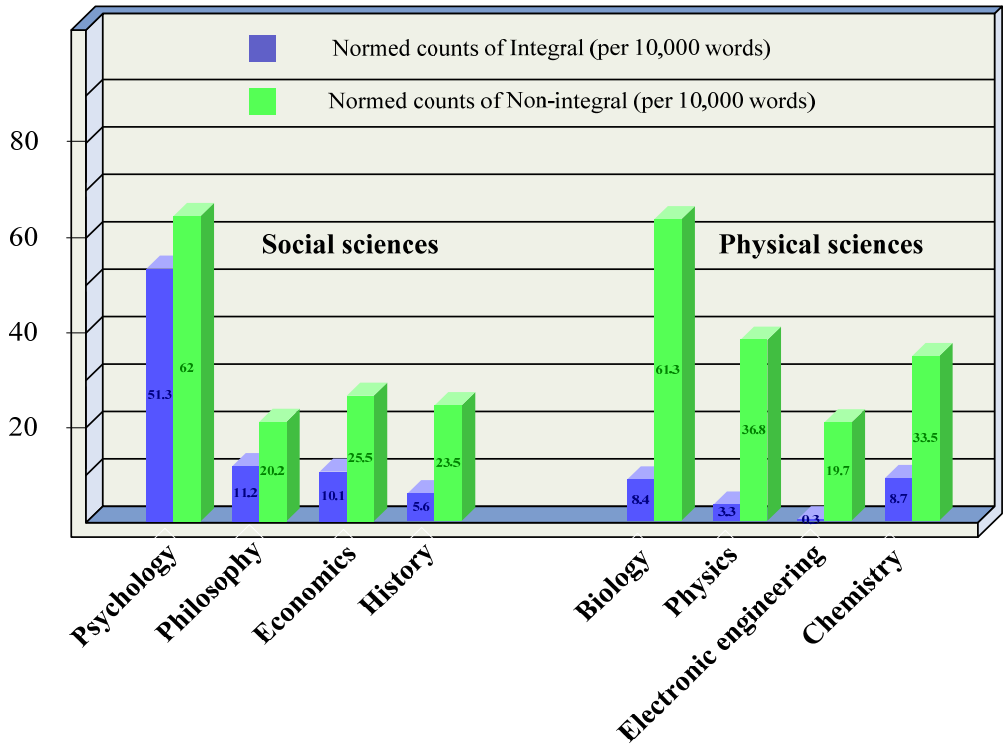

Figure 4.4 shows frequency of integral-ness and prominence in different disciplines in social sciences and physical sciences

Figure 4.4 depicts variation in citation structure in different disciplines. As seen from Figure 4.4, within the same discipline, there are more non-integral citations than integral citations. In social sciences, Psychology and Philosophy use integral structures most frequently. Except Psychology, other disciplines use nearly the same amount of non-integral citations per 1000 words. In physical sciences, Physics and Electronic engineering use the least integral citations.

Although in social sciences, Psychology and Philosophy use the most integral citations, there are more non-integral citations than integral citations in Psychology and Philosophy respectively. This result is inconsistent with Hyland's (1999) results. Hyland's (1999) study shows Philosophy is the only discipline that prefers integral structure to non-integral structure. By examining the 10 theses from Philosophy, it is found that five theses are argument or analysis to a theory of a specific philosopher. The possibly plausible explanation is that writers in Philosophy in sample tend to give prominence to theory of the cited author and it take for granted that all the argument and analysis are aimed at the theories of the cited author, therefore information about theory is given more emphasis.

Table 4.6 Proportion of integral and non-integral citations according to language and area

Functions	Chinese social sciences		Chinese physical sciences		English social sciences		English physical sciences	
	Raw count	%	Raw count	%	Raw count	%	Raw count	%
Integral citations	174	45.9%	87	19.0%	162	32.3%	35	7.2%
Non-integral citations	205	54.1%	370	81.0%	339	67.7%	451	92.8%
Total	379	100%	457	100%	501	100%	486	100%

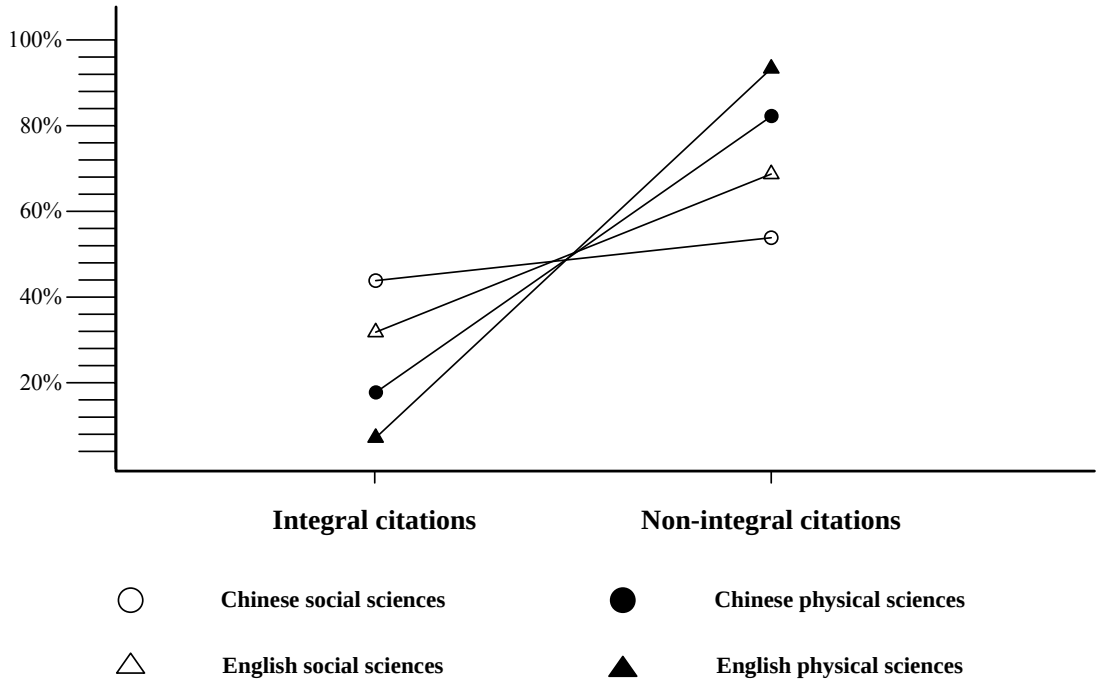

Figure 4.5 shows proportion of integral and non-integral citations according to language and area

Figure 4.5 shows distribution of integral and non-integral citations in four areas: Chinese social sciences, Chinese physical sciences, English social sciences and English physical sciences. As shown in Table 4.6, more non-integral citations than integral citations are used in these four fields. In terms of integral citations, the proportion of Chinese social sciences is the highest while that of English physical

sciences is the lowest. In regard to non-integral citations, English physical sciences data share the highest percentage with Chinese physical sciences and English social sciences following the second and the third respectively.

4.2 Reporting forms

In this study, forms of reporting in the sample are discussed in terms of five categories: short direct quotes, extensive direct quotes, summary, generalization and list for reference. Based on the categories of Hyland (1999), list for reference is added in the present study.

Table 4.7 Frequency of five reporting forms in Chinese corpus and English corpus

Forms	Corpus of Chinese theses			Corpus of English theses		
	Raw counts	Percentage	Normed counts (per 10000 words)	Raw counts	Percentage	Normed counts (per 10000 words)
Short direct quotes	2	0.2%	0.1	31	2.6%	2.6
Extensive direct quotes	41	4.2%	1.8	62	5.2%	5.3
Summary	791	80.0%	34.6	881	73.5%	74.6
Generalization	144	14.6%	6.3	172	14.4%	14.6
List for reference	11	1.1%	0.5	52	4.3%	4.4

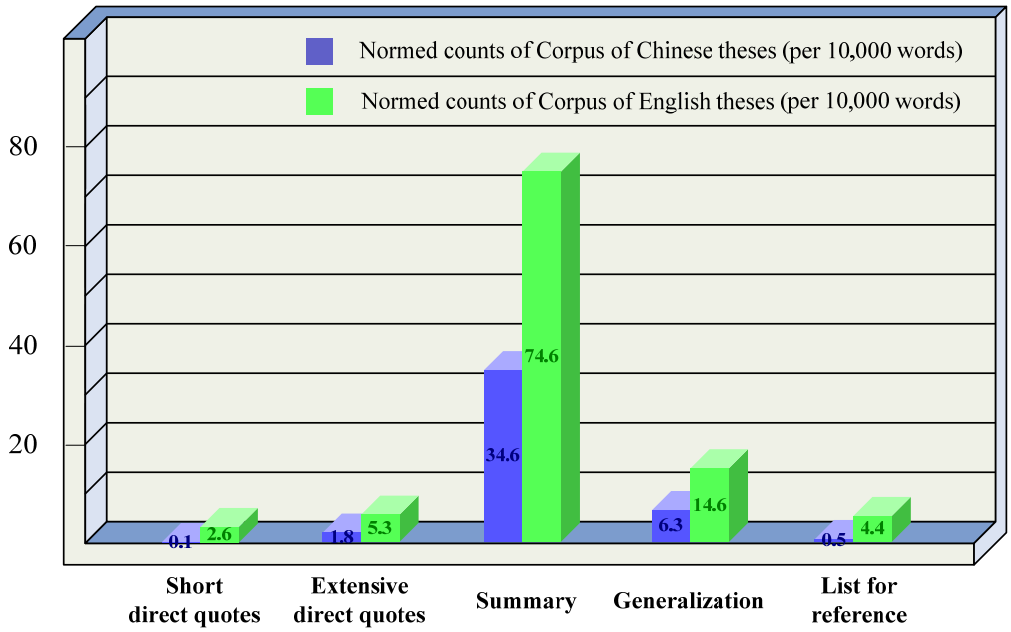

Figure 4.6 shows frequency of five reporting forms in Chinese corpus and English corpus

As shown in Table 4.7, in both Chinese corpus and English corpus, summary is used most frequently (80.0% and 89.1% respectively), with generalization ranking the second. This result is consistent with the findings of Dubois (1988) and Hyland (1999). Short direct quotes are least used (0.2% and 2.6% respectively) in the two corpora. According to Hyland (1999), the way to present reported information is crucial in gaining acceptance for a claim the writer makes, so the writer tends to choose the most effective ways to support their own argument. 'Summary and generalization are the most effective ways of achieving this as they allow the writer greater flexibility to emphasize and interpret what they are citing' (p.348).

It is seen from Figure 4.6 that there are more normed counts of every category per 10000 words in the English corpus than in the Chinese corpus. The difference in list for reference between two corpora is most noticeable. Abbreviations like e.g./cf./see often occur in English articles to indicate further information for readers to make reference to related literature, to make comparison of some argument, ideas or conceptions between different sources. However, in Chinese writing, writers seldom have this habit to provide relevant information.

Table 4.8 Frequency of five reporting forms
in social sciences and physical sciences

Forms	Data of Social sciences		Data of Physical sciences	
	Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Short direct quotes	29	1.9	4	0.2
Extensive direct quotes	99	6.5	4	0.2
Summary	892	58.6	780	40.2
Generalization	56	3.5	260	13.4
List for reference	53	3.5	10	0.5

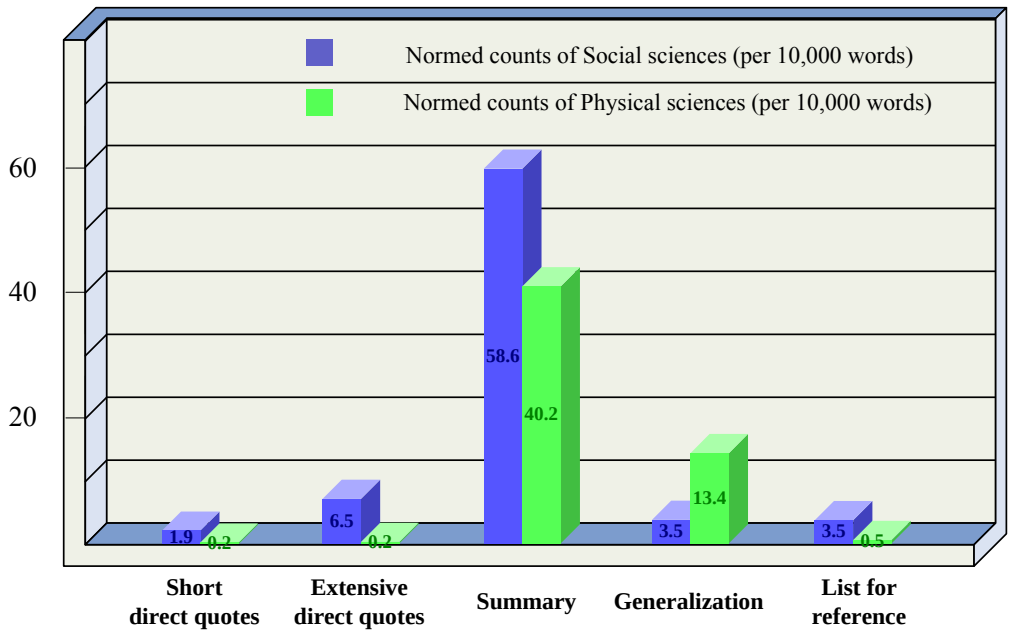

Figure 4.7 shows frequency of five reporting forms
in social sciences and physical sciences

It can be seen from Figure 4.7 that in both social sciences and physical sciences most reporting statements are expressed in the form of summary; however, in social sciences extensive direct quotes rank the second while in physical sciences generalization comprises most of the rest. Except generalization, short direct quotes, extensive direct quotes, summary and list for reference are used more frequently in social sciences than in physical sciences. There is a greatest distance in direct quotes

(including short and extensive ones) between social sciences and physical sciences. The possible reasons are explored below.

Table 4.9 Frequency of five reporting forms in different disciplines in social sciences and physical sciences

Disciplines	Short direct quotes		Extensive direct quotes		Summary		Generalization		List for reference	
	RC	NC	RC	NC	RC	NC	RC	NC	RC	NC
Psychology	16	3.4	23	4.9	521	9.9	35	7.4	40	8.4
Philosophy	1	0.4	34	14.6	63	27.0	1	0.4	1	0.4
Economics	9	1.9	26	5.5	193	40.7	11	2.3	4	0.8
History	3	0.9	16	4.7	115	33.7	9	2.6	8	2.3
Biology	4	0.5	1	0.1	408	55.0	149	20.1	10	1.3
Physics	0	0	3	1.0	113	37.1	26	8.5	0	0
Electronic engineering	0	0	0	0	63	18.8	16	4.8	0	0
Chemistry	0	0	0	0	196	34.9	69	12.3	0	0

(Note: RC stands for raw counts; NC stands for normed counts per 10000 words)

As shown in Table 4.9, direct quotes occur most frequently in Philosophy, with Psychology and Economics following the second and the third. In Philosophy, writers tend to quote the original words of the cited authors directly and then propose agreement, disagreement or analysis of the ideas of the cited authors. Moreover, direct quotes of the author's original words are the most vivid and direct way to present the cited author's ideas. Observed from Table 4.9, in physical sciences, only summary and generalization are adopted in Electronic engineering and Chemistry. There is no count of other reporting forms in Electronic engineering and Chemistry. It may be easy to conclude that writers in social sciences tend to use more forms of reporting in thesis writing than writers in physical sciences.

Table 4.10 Proportion of reporting forms according to language and area

Forms	Chinese social sciences		Chinese physical sciences		English social sciences		English physical sciences	
	Raw count	%	Raw count	%	Raw count	%	Raw count	%
Short direct quotes	2	0.3%	0	0%	27	4.4%	4	0.7%
Extensive direct quotes	41	8.0%	0	0%	58	9.4%	4	0.7%
Summary	424	82.5%	367	77.3%	468	76.1%	413	70.8%
Generalization	38	7.4%	106	22.3%	18	2.9%	154	26.4%
List for reference	9	1.8%	2	0.4%	44	7.2%	8	1.4%
Total	514	100%	475	100%	615	100%	583	100%

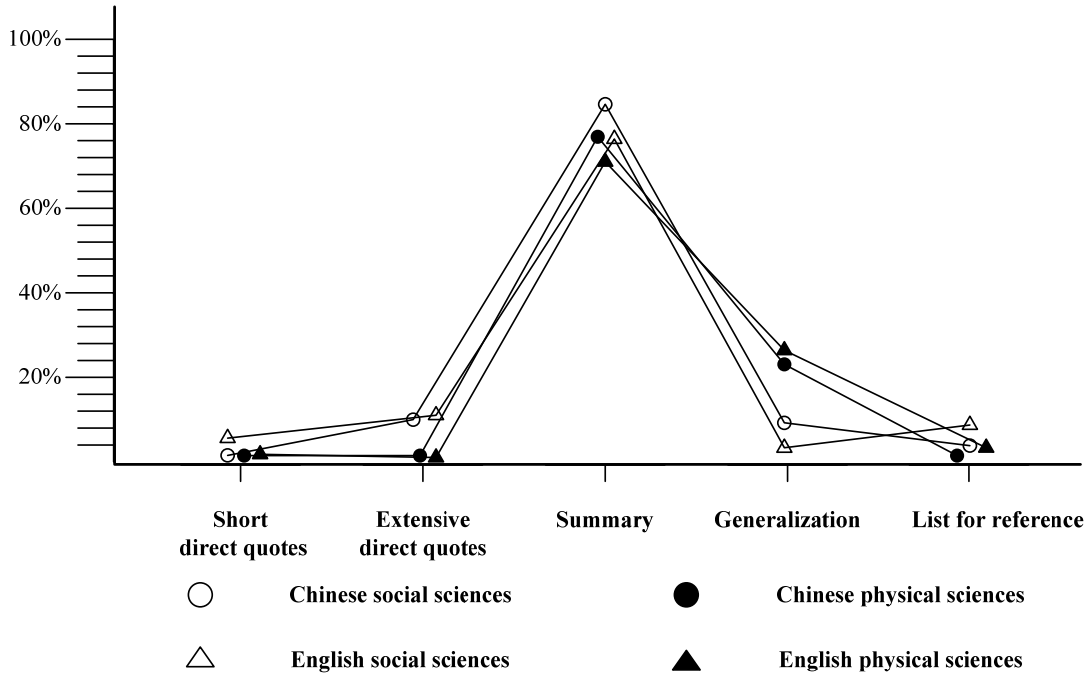

Figure 4.8 shows proportion of five reporting forms according to language and area

Seen from Figure 4.8, summary is used most frequently in these four fields while short direct quotes are least used. No direct quotes are found in Chinese physical sciences. There are more generalizations in physical sciences in both Chinese and English than in social sciences in both Chinese and English.

Table 4.11 Proportion of each reporting form to total forms

Forms	Raw counts	Proportion
Short direct quotes	33	1.5%
Extensive direct quotes	103	4.7%
Summary	1672	76.5%
Generalization	316	14.4%
List for reference	63	2.9%
Total	2187	100%

In total reporting forms of both Chinese data and English data, summary is used most frequently and generalization ranks the second.There are more extensive direct quotes than short direct quotes which are least used.

4.3 Reporting verbs

4.3.1 Categories of reporting verbs

This section investigates categories of reporting verbs based on the classification of Thompson and Ye (1991) in terms of denotation and evaluation. First, denotation of reporting verb is examined. Although Thompson and Ye's (1991) categories of denotation is proposed to analyze reporting verbs academic papers in English, it is found in the sample their classification can also be applied to academic papers in Chinese.

Table 4.12 Frequency of reporting verbs in denotation in Chinese corpus and English corpus

Categories of denotation		Corpus of Chinese theses		Corpus of English theses	
		Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Author acts	Textual	144	6.3	164	13.9
	Mental	36	1.6	8	0.7
	Research	78	3.4	84	7.1
Writer acts	Comparing	0	0	0	0
	Theorizing	0	0	3	0.3

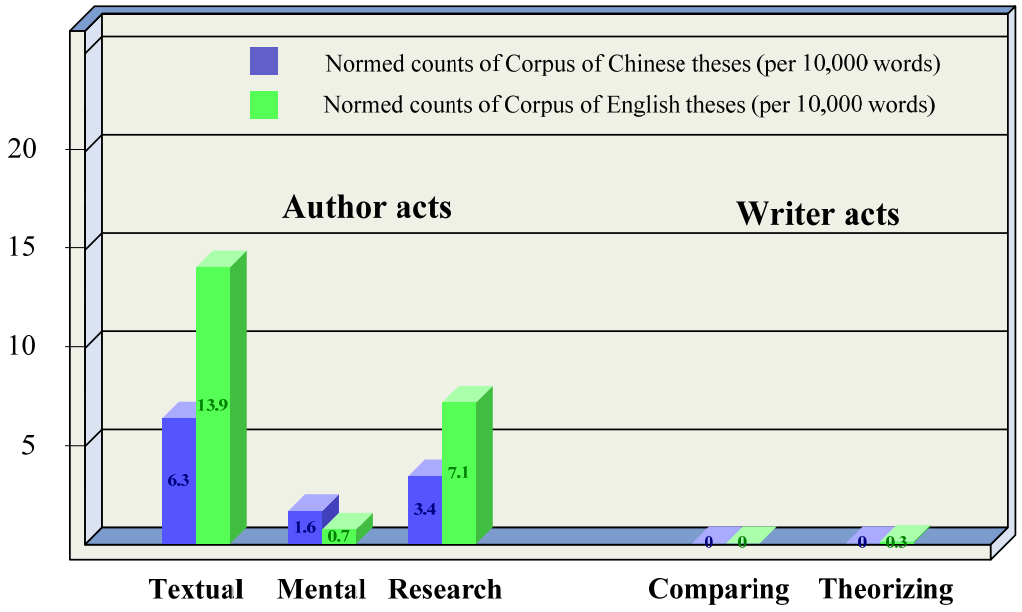

Figure 4.9 shows frequency of reporting verbs in denotation in Chinese corpus and English corpus

Generally speaking, it is found that writers in English corpus use more reporting verbs than writers in Chinese corpus. Writers in English corpus use more textual and research reporting verbs than writers in Chinese corpus; however, mental verbs are used more frequently in Chinese corpus than in English corpus. Writers in Chinese corpus tend to use mental verb RENWEI (it is equivalent to *think* in English) more frequently. Comparing and theorizing verbs are hardly found in either of the two

corpora. There are no comparing verbs found in two corpora, and only three theorizing verbs *support (occurring two times)*, *account for* are found in English corpus. For example,

> *In a similar manner, the lack of lexical bias in Del Viso et al.'s (1991) study could be accounted for by the fact that Spanish contains a greater number of longer words than English (Berg, 1991)...*

> *Motor schema theory (Schmidt, 1975) helps to understand the movement that occurs during web browsing. It supports a generalized motor program, controlling movement exwcution...*

Next, denotation of reporting verb in social sciences and physical sciences is analyzed.

Table 4.13 Frequency of reporting verbs in denotation in social sciences and physical sciences

Categories of denotation		Social sciences		Physical sciences	
		Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Author acts	Textual	226	14.8	82	4.2
	Mental	44	2.9	0	0
	Research	99	6.5	63	3.2
Writer acts	Comparing	0	0	0	0
	Theorizing	3	0.2	0	0

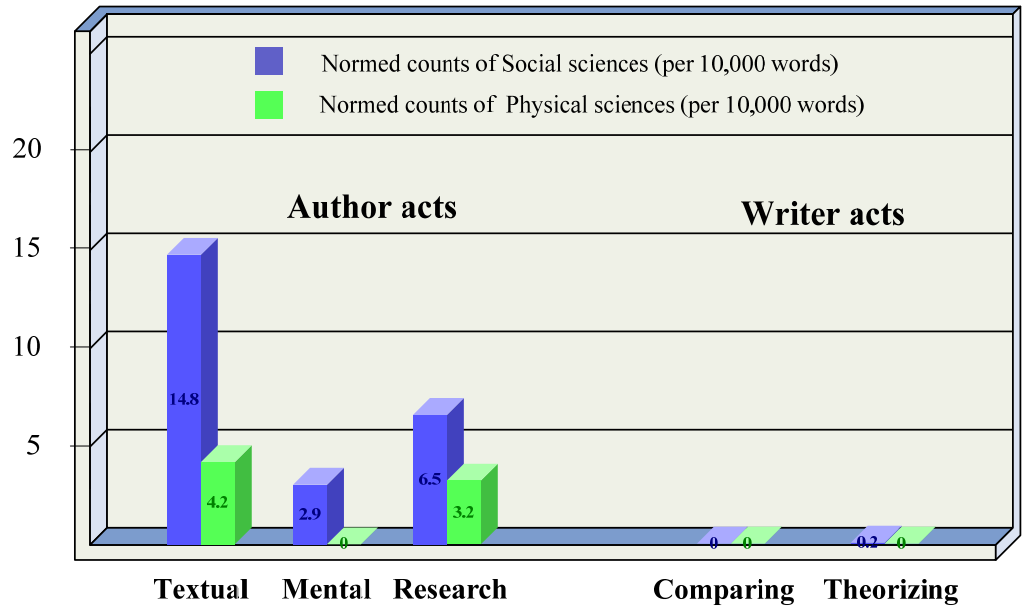

Figure 4.10 shows frequency of reporting verbs in denotation in social sciences and physical sciences

As show in Figure 4.10, reporting verbs in each category of denotation occur much more frequently in social sciences than in physical sciences. There are 14.8 textual verbs per 10000 words in social sciences while 4.2 per 10000 words in physical sciences. This result is consistent with Hyland's (1999) finding that social sciences such as philosophy, sociology, marketing favoured textual reporting verbs; however, Figure 4.10 shows that social sciences use more research verbs than physical sciences, which is not consistent with Hyland's (1999) conclusion that engineering and science papers favoured research reporting verbs. The reason of this inconsistence between present study and previous study may lie in the discipline of Psychology. An overwhelming majority of research reporting verbs occur in Psychology. By examining all the theses of Psychology in sample, it is found these papers are all empirical studies based on methods of experiment in which research verbs often occur. Theses of the other disciplines such as history, philosophy etc. are usually theoretical studies.

It is interesting to find that there is no use of mental verbs in the physical sciences. Social sciences to some extent depends on individual understanding and individual interpretation while discipline in physical sciences is pure science which is more objective. Mental reporting verbs which describe the mental processes of the cited

authors are chosen from the perspectives of the writers, although mental processes are ascribed to the cited authors. Mental verbs such as *focus on, believe, consider* etc. are to some extent subjective, which depending on writers' purpose of citing the author.

Table 4.14 Proportion of denotational categories of reporting verbs according to language and area

Categories	Chinese social sciences		Chinese physical sciences		English social sciences		English physical sciences	
	Raw count	%	Raw count	%	Raw count	%	Raw count	%
Textual	112	61.5%	32	42.1%	114	61.0%	50	69.4%
Mental	36	19.8%	0	0%	8	4.2%	0	0%
Research	34	18.7%	44	57.9%	65	34.8%	19	26.4%
Comparing	0	0%	0	0%	0	0%	0	0%
Theorizing	0	0%	0	0%	0	0%	3	4.2%
Total	182	100%	76	100%	187	100%	72	100%

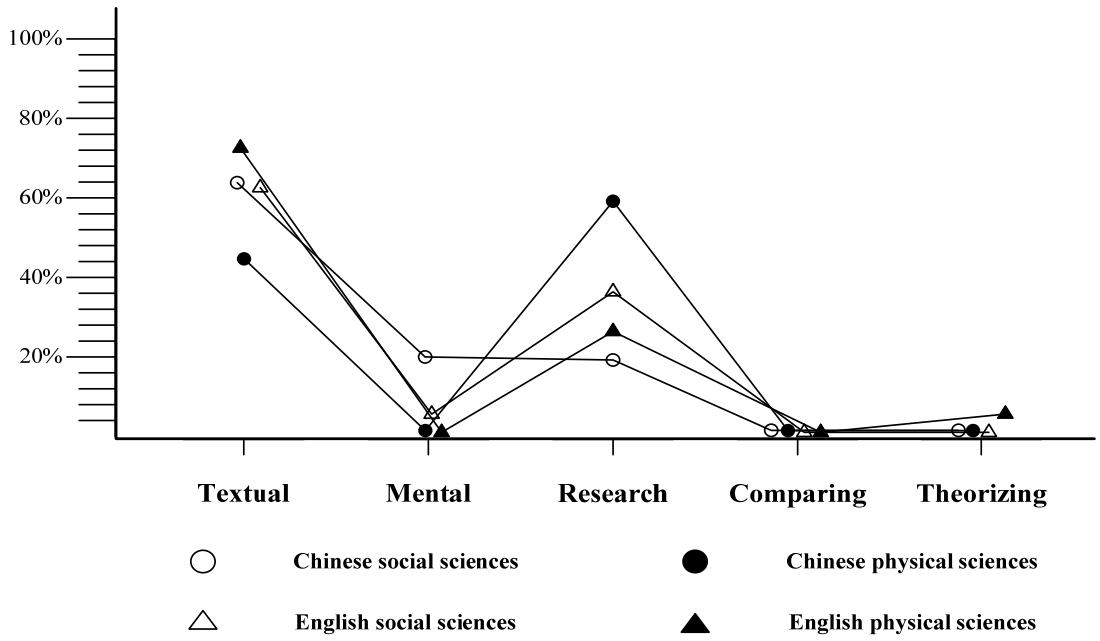

Figure 4.11 shows distribution of denotational categories of reporting verbs according to language and area

As shown in Figure 4.11, textual verbs in four fields share a highest percentage with research verbs following the second. There are fewer textual verbs in Chinese physical sciences and the percentages in other three fields are close. Comparing and theorizing verbs are rarely found in four fields. More mental verbs in Chinese social sciences are used than in other three fields.

Table 4.15 Proportion of denotational categories of reporting verbs to all reporting verbs

Categories	Raw counts	Proportion
Textual	308	59.6%
Mental	44	8.5%
Research	162	31.3%
Comparing	0	0
Theorizing	3	0.6%
Total	517	100%

It is found that textual verbs have a highest percentage with research verbs ranking the second. There are no comparing verbs and only 0.6% theorizing verbs occur.

Then we come to discussion of evaluation in reporting verbs. Three factors author's stance, writer's stance and writer's interpretation are involved in analyzing the evaluation of reporting verbs (Thompson & Ye 1991). In this study, we focus on the author's stance and writer's stance. It is found that reporting verbs in corpus of Chinese theses can also be classified according to author's stance and writer's stance.

Table 4.16 Frequency of reporting verbs in evaluation in Chinese corpus and English corpus

Categories		Corpus of Chinese theses		Corpus of English theses	
		Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Author's stance	positive	64	2.8	65	5.5
	Negative	1	0.04	8	0.7
	Neutral	23	1.0	49	4.2
Writer's stance	Factive	67	2.9	46	3.9
	Counter-factive	0	0	2	0.2
	Non-factive	116	5.1	127	10.8

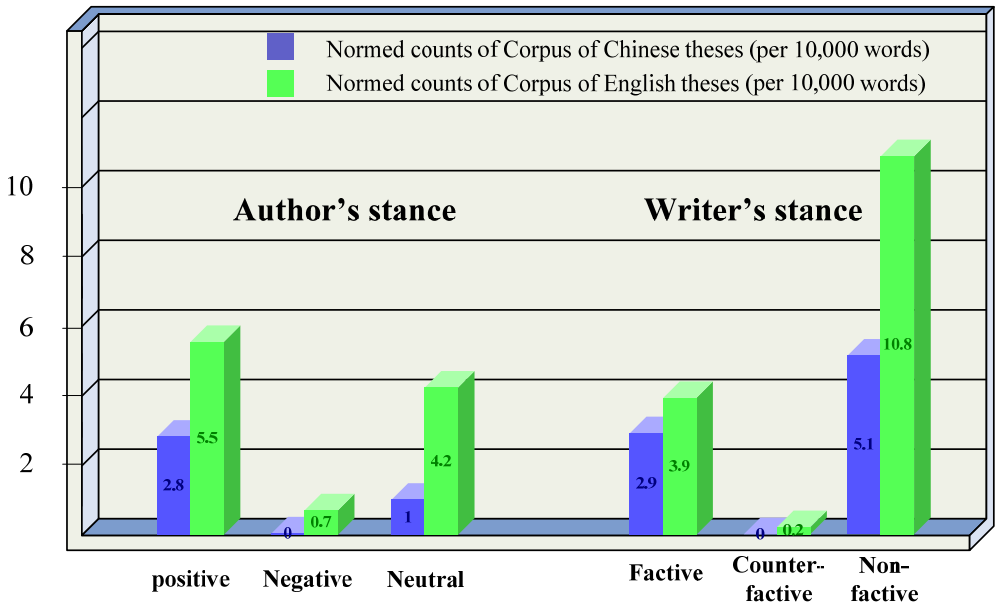

Figure 4.12 shows frequency of reporting verbs in evaluation in Chinese corpus and English corpus

Generally speaking, there are more counts per 10000 words of each category of evaluation in reporting verbs in English corpus than in Chinese corpus. Figure 4.12 shows that writers in English corpus use about twice as many positive reporting verbs as writers in Chinese corpus. It is the same case with use of non-factive reporting verbs in two corpora. It is also found that in both corpora, in author's stance, negative

reporting verbs are used most rarely, while in writer's stance counter-factive ones are used most rarely. It is observed that no counter-factive reporting verb is used in Chinese corpus. The biggest difference between the two corpora lies in the amount of negative reporting verbs used in two corpora: 0.7 per 10000 words are found in English corpus while 0.04 per 10000 words found in Chinese corpus.

Table 4.17 Frequency of reporting verbs in evaluation in social sciences and physical sciences

Categories of evaluation		Social sciences		Physical sciences	
		Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Author's stance	positive	92	6.0	37	1.9
	Negative	7	0.5	2	0.1
	Neutral	58	3.8	14	0.7
Writer's stance	Factive	61	4.0	52	2.7
	Counter-factive	2	0.1	0	0
	Non-factive	187	12.3	56	2.9

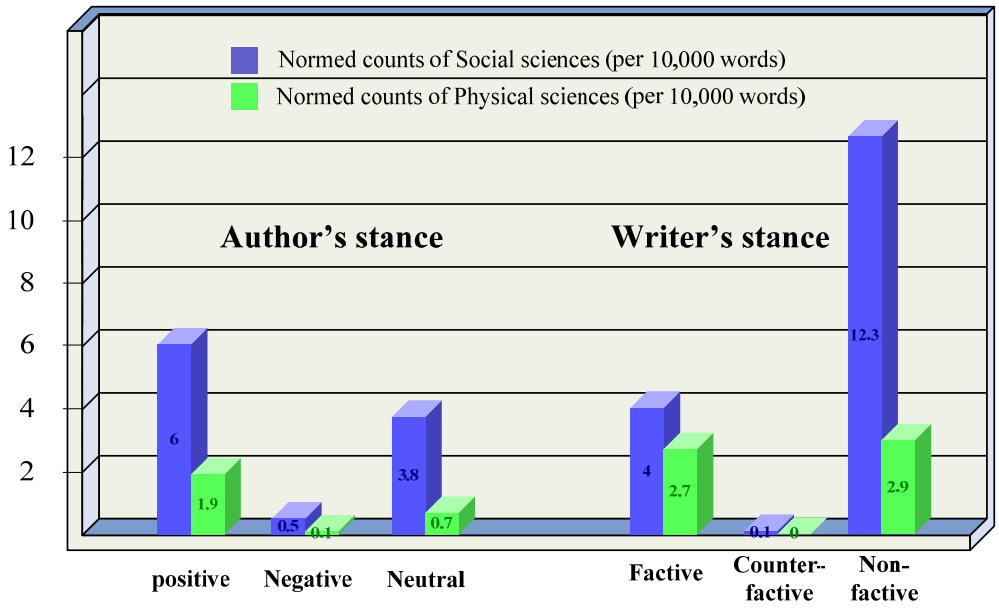

Figure 4.13 Frequency of reporting verbs in evaluation in social sciences and physical sciences

On the whole, much higher counts (per 10000 words) of each category of evaluation in reporting verbs are used in social sciences than physical sciences. In author's stance, positive reporting verbs are used most frequently in both corpora while negative verbs are most rarely used in both corpora. In writer's stance, non-factive reporting verbs are used more frequently than factive verbs in both corpora. By examining evaluative verbs in each discipline, it is found that the amount of factive verbs is still exceeded by that of non-factive verbs in all disciplines. Only two counter-factive verbs are found in social sciences.

Table 4.18 Proportion of evaluative categories of reporting verbs according to language and area

Categories	Chinese social sciences		Chinese physical sciences		English social sciences		English physical sciences	
	Raw count	%	Raw count	%	Raw count	%	Raw count	%
Positive	43	23.2%	21	24.4%	49	22.1%	16	21.3%
Negative	0	0%	1	1.2%	7	3.2%	1	1.3%
Neutral	16	8.6%	7	8.1%	42	18.9%	7	9.3%
Factive	20	10.8%	47	54.7%	41	18.5%	5	6.8%
Counter-factive	0	0%	0	0%	2	0.9%	0	0%
Non-factive	106	57.3%	10	11.6%	81	36.5%	46	61.3%
Total	185	100%	86	100%	222	100%	75	100%

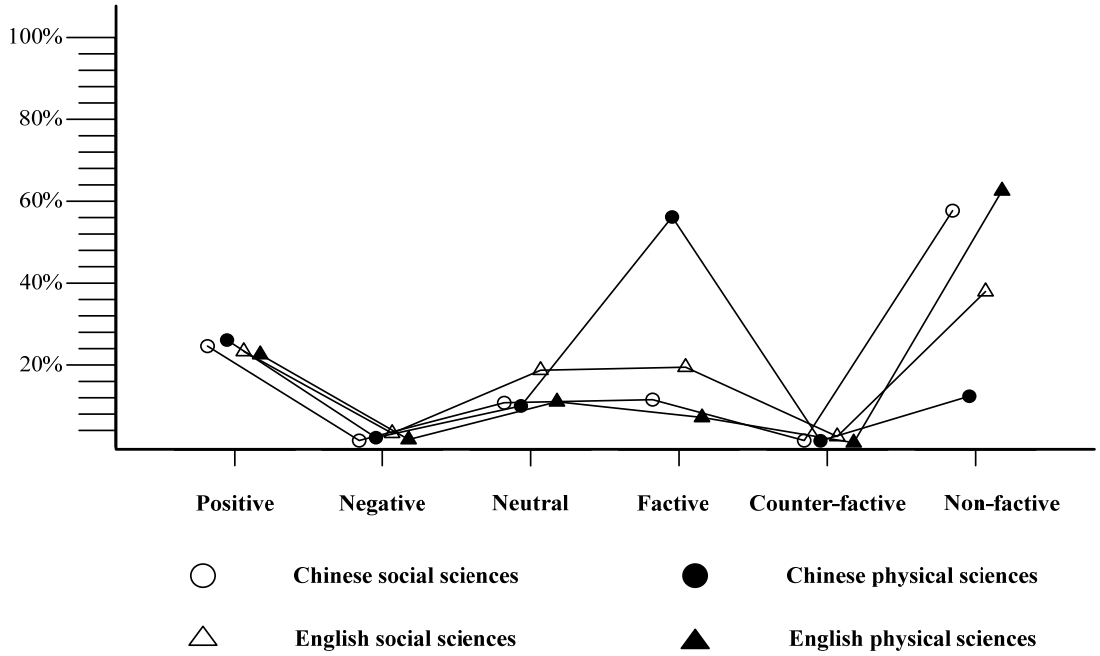

Figure 4.14 shows distribution of evaluative categories of reporting verbs according to language and area

In author's stance, positive verbs are used most frequently while negative are least used. Percentages of positive verbs in four fields are close and it is the same case with negative and neutral verbs. In writer's stance, non-factive verbs are used most frequently on the whole while counter-factive verbs occur rarely in four fields. More factive verbs are used in Chinese physical sciences than in other fields while fewer non-facitve verbs are used in Chinese physical sciences.

Table 4.19 Proportion of evaluative categories of reporting verbs to all reporting verbs

Categories	Raw counts	Proportion
Positive	129	22.7%
Negative	9	1.6%
Neutral	72	12.7%
Factive	113	19.9%
Counter-factive	2	0.3%
Non-factive	243	42.8%
Total	568	100%

Non-factive verbs are used most frequently and positive and factive verbs follow the second and the third. Negative and counter-factive verbs are rarely chosen, only 1.6% and 0.3% respectively.

4.3.2 Tense of reporting verbs

In Chinese (here Chinese refers to Mandarin), tense is usually expressed by lexical clues such as time expressions. In Chinese, verbs don't inflect or change its form to indicate tense; however, the verbs are linked with an aspect marker which may indicate the completion, experience, continuation, etc. of an action (Yip & Rimmington 1998). The most common aspect marker found in sample is *LE* which indicates completed or past action. LE usually functions as a verb suffix which follows the verb. In the 'V+LE' construction, LE stresses the completion of an action at a specific time in the past, present, or future (Lin 1984). The analysis of tense and aspect of reporting verbs in Chinese data is more complex than those in English data. If a Chinese sentence without lexical clues such as time expressions is divorced from context, we cannot judge tense of verb in this sentence. Time expressions set the time context for the action of the verb (Yip & Rimmington 1998). There are several cases : one is a time expression in a sentence can indicate the past time of verb explicitly (see *in 1988* in example 1); the second one is that there are no clear clues or signals in the sentence to indicate whether the verb (without LE following) expresses present time or past time (see examples 2 and 3); the third one is the verb is considered as referring to past time because the time expression can clearly indicate the action occurred in the past (see *once* in example 4); the last one is 'verb + LE' construction in a sentence without other clues to express tense can indicate either present perfect or past tense (see example 5). Consider the following examples from the sample:

1) **1998 年**，丹尼尔.戈尔曼在他的《EQⅡ——工作 EQ》一书中**提出了**工作 EQ 的模型。

   [Phonetic transcription: **1998nian**, Daniel Goleman zai tade 《EQⅡ——gongzuoEQ》yishu zhong **tichuLE** EQdemoxing]

(Literal translation: 1998 year, Daniel Goleman in his book *EQ Ⅱ---Working EQ* proposed model of Working EQ.)

English translation: **In 1998**, Daniel Goleman **proposed** the model of Working EQ in his book *EQ Ⅱ---Working EQ*.

2) Barrera 和 Alinlay （1983）**将**社会支持**分为**六类。
[Phonetic transcription: Barrera he Alinlay (1983) **jiang** shehui zhichi **fenwei** liulei]

（Literal translation: Barrera and Alinlay classify social support six categories.）

English translation: Barrera and Alinlay (1983) **classify/classified** social support into six categories.

3) Bettman, Luce & Payne (1998) **认为**信息加工策略是方案数目的函数
[Phonetic transcription: Bettman, Luce & Payne (1998) **renwe**i xinxi jiagong celue shi fang'an shumu de hanshu]

(Literal translation: Bettman, Luce & Payne (1998) think Information Processing Strategy is number of schemes's function.)

English translation: Bettman, Luce & Payne (1998) **think/thought** Information Processing Strategy is the function of the number of schemes.

The reporting verbs *classify* in example 2) and *think* in example 3) can be regarded as expressing past time or present time because the tense of this sentence cannot be judged from this sentence itself or the co-text. Considerable examples like this can be found in the Chinese data.

4) 著名社会学家韦伯**曾论述了**两类信任：普遍信任和特殊信任。[Phonetic transcription: zhuming shehuixuejia weibo ceng lunshuLE lianglei xinren: pubian xinren he teshu xinren.]

（Literal translation: famous sociologist Wei Bo **once discuss** two kinds trust: common trust and special trust.）

English translation: The famous sociologist Wei Bo **once discussed** two kinds of trust: common trust and special trust.

The adverbial *once* in example 4) is a marker which indicates the action 'discuss' in this sentence happened in the past.

5) Beatty 和 Smith （1987） **提出了**这样的消费行为模型： 行为=个体+任务+（个体×任务）+ 误差。 [Phonetic transcription: Beatty he Smith （1987）**tichuLE** zheyangde xiaofei xingwei moxing: xingwei = geti + renwu + (geti×renwu) + wucha.]

(Literal translation: Beatty and Smith （1987）propose such Purchase Behavior Model: behavior = individual + task + (individual×task) + error.)

English translation: Beatty and Smith （1987）**proposed/have proposed** such a Purchase Behavior Model: behavior = individual + task + (individual×task) + error.)

In contrast with Chinese, tense in English finds expression in morphological change of verbs. Present tense, past tense and present perfect are found in English corpus. (Here present, past and present perfect are regarded as three tenses because I'm not concerned with the distinction between tense and aspect.)

Table 4.20 Frequency of tense in Chinese corpus and English corpus

Tense	Corpus of Chinese theses		Corpus of English theses	
	Raw counts	Normed counts (per 10000 words)	Raw counts	Normed counts (per 10000 words)
Present			127	10.8
Past	76	3.3	126	10.7
Present perfect			9	0.8
Indefinite	408	17.9		

Table 4.20 shows that there are almost the same amount of present tense and past tense used in English corpus, with present perfect occupying an extremely small percentage. The tense usage of reporting verbs are nearly the same as the previous studies (Oster, 1981; Malcolm 1987; Weissberg and Buker, 1990) indicate. Past tense occurs in a reporting sentence referring to a specific study or experiment whose findings are only limited to the cited study or experiment. Present tense is used in the findings which are believed as fact by the writer or supportive of the current study. Present perfect often occurs in general statements of research activity as a whole in an area. In contrast with English corpus, only about 3.3 reporting verbs which mark past tense per 10000 words can be definitely identified in Chinese corpus, and the rest reporting verbs can be regarded as present or past tense, and past tense or present perfect. In English, difference between 'report of specific study' and 'generally accepted fact' is generally conveyed by past tense and present tense respectively; however, in Chinese reporting verbs cannot convey this difference clearly. In most cases, 'report of specific study' in Chinese is expressed in past time which is indicated by time expressions or adverbials. 'Report of specific study' usually describes a process or procedure of a study or an experiment by using research verbs. It is difficult to judge whether a reporting statement is accepted as fact or not. This can be inferred from the context rather than the expression of past time or present time.

The tense of the majority of reporting verbs in Chinese corpus is indefinite; therefore, the comparison between social science and physical science is not discussed here.

4.4 Function of reporting

In this study, Bloch and Chi's (1995) functions of *background, support, faulty path and return path* are used. Bloch and Chi (1995) point out it is often difficult to distinguish exactly whether a citation is being used as a faulty path where the author disagrees with the citation, or return path where the citation is used to support a point the author disagrees with. Bloch and Chi uses a category called critical to total faulty path and return path together. Three functions of reporting are examined in the sample: background, support and critical.

Table 4.21 Frequency of function in Chinese corpus and English corpus

Functions	Corpus of Chinese theses			Corpus of English theses		
	Raw counts	Normed counts (per 10000 words)	Percentage	Raw counts	Normed counts (per 10000 words)	Percentage
Background	982	43.0	95.9%	988	83.7	86.6%
Support	27	1.2	2.6%	45	3.8	3.9%
Critical	15	0.7	1.5%	108	9.2	9.5%

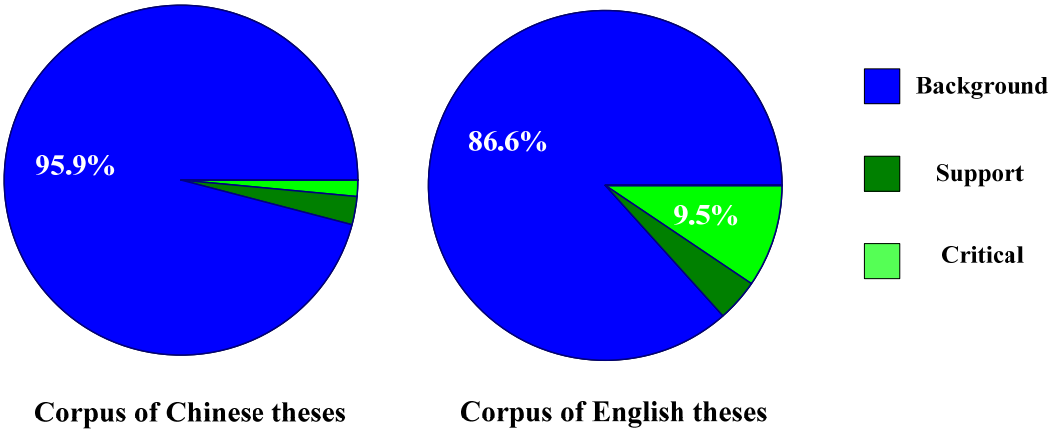

Figure 4.15 Percentages of three functions
in Chinese corpus and English corpus

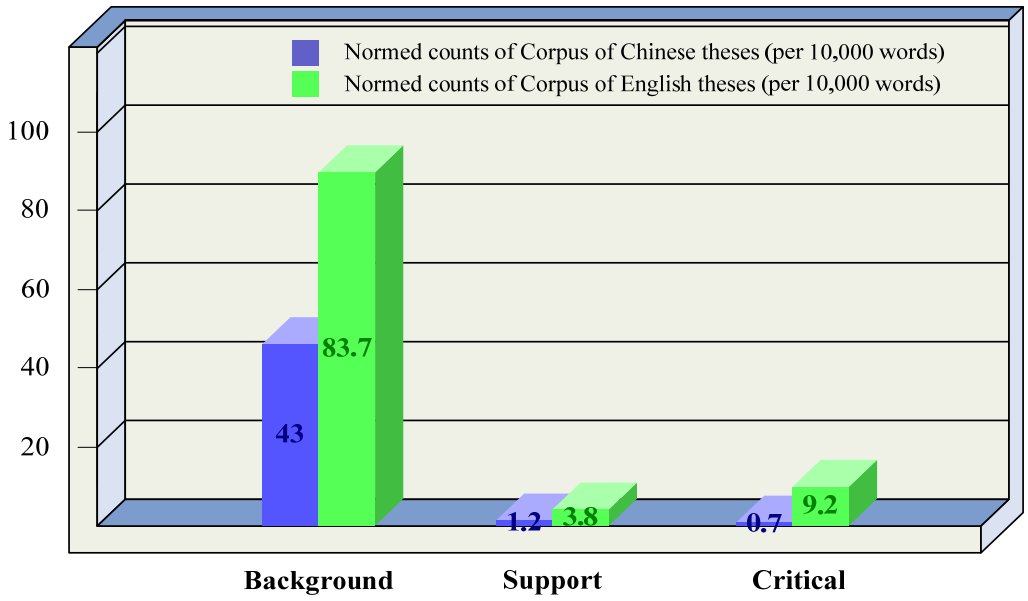

Figure 4.16 shows normed counts of three functions
in Chinese corpus and English corpus

As shown in Figure 4.15, in Chinese corpus function of background is used most frequently, with function of support ranking the second and critical function the third. In English corpus, most reporting sentences are giving background (about 86.6%), critical reporting occupies 9.5% and the function of support is used least. It can be seen from Figure 4.16, overwhelming majority reporting sentences in both corpora are used to give background; however, more there are more critical reporting in English corpus than in Chinese corpus. As normed counts per 10000 words show, there are more reporting in English corpus than in Chinese corpus in each function, especially critical function.

Table 4.22 Frequency of tense in social sciences and physical sciences

Functions	Social sciences			Physical sciences		
	Raw counts	Normed counts (per 10000 words)	Percentage	Raw counts	Normed counts (per 10000 words)	Percentage
Background	948	62.3	87.4%	1022	52.6	94.6%
Support	53	3.5	4.9%	19	1.0	1.8%
Critical	84	5.5	7.7%	39	2.0	3.6%

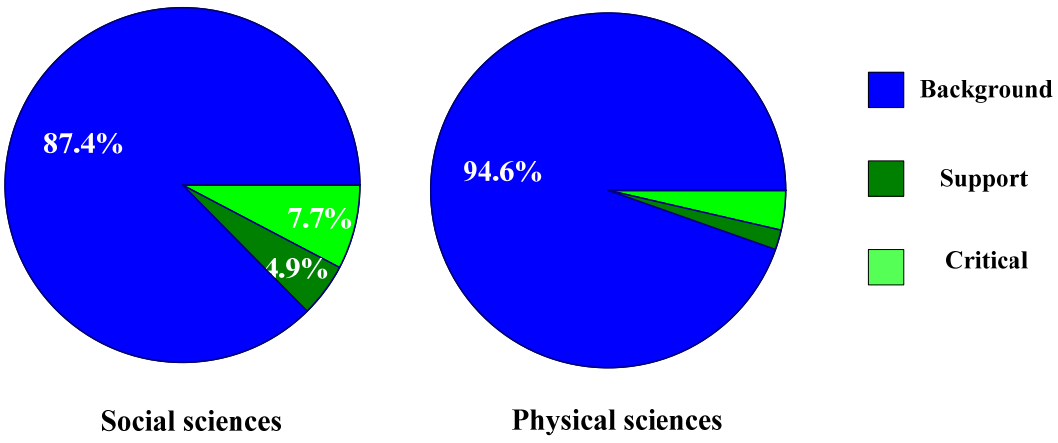

Figure 4.17 Percentages of three functions in social sciences and physical sciences

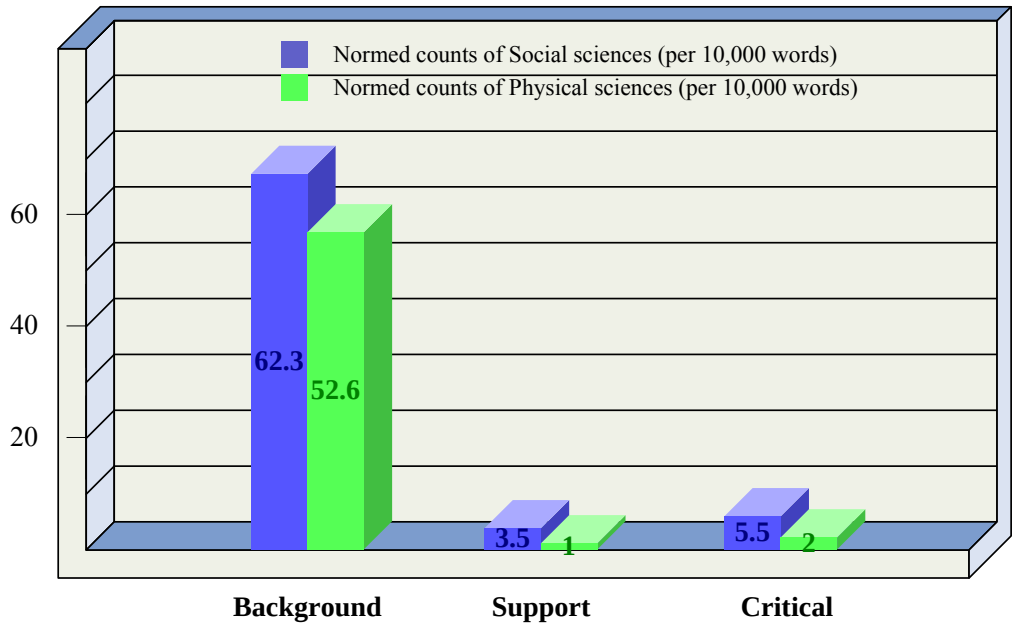

Figure 4.18 Normed counts of three functions
in social sciences and physical sciences

Seen from Figure 4.17, a great number of reporting statements are used to give background information (87.4% and 94.6%), and the least is used to give support to an argument in both social sciences and physical sciences. Figure 4.18 shows that more reporting per 10000 words in each function is used in social sciences than in physical sciences. The difference of background between social and physical sciences is not so distinct while the differences of support and critical are more noticeable.

Next, I shall investigate functional differences in reporting in the sample according to language and area.

Table 4.23 Percentages of reporting functions according to language and area

Functions	Chinese social sciences		Chinese physical sciences		English social sciences		English physical sciences	
	Raw count	%	Raw count	%	Raw count	%	Raw count	%
Background	500	93.5%	482	98.6%	448	81.5%	540	91.4%
Support	21	3.9%	6	1.2%	32	5.8%	13	2.2%
Critical	14	2.6%	1	0.2%	70	12.7%	38	6.4%
Total	535	100%	489	100%	550	100%	591	100%

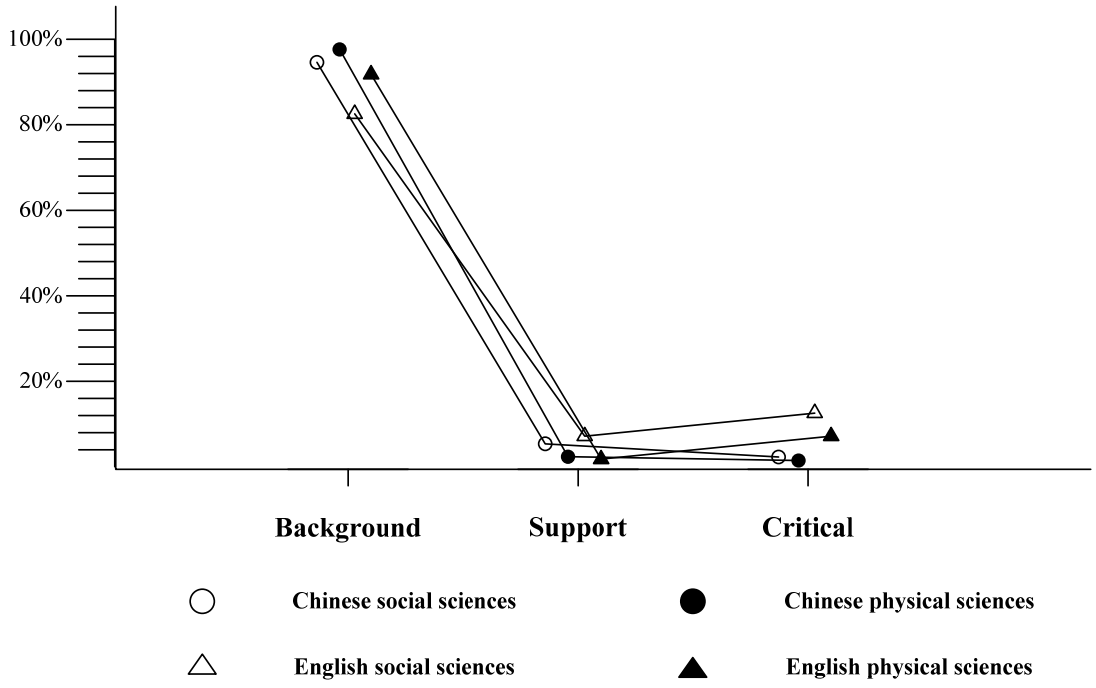

Figure 4.19 shows the distribution of percentages of reporting functions according to language and area

Figure 4.19 shows a distribution of percentages of reporting functions of giving background, providing support and expressing critical views in four areas: Chinese social sciences, Chinese physical sciences, English social sciences and English physical sciences. As shown in Figure 4.19, function of giving background shares the highest percentage of three functions. Writers in Chinese physical sciences use the

highest percentage of function of giving background while English social sciences use a lowest percentage. It is found that in both Chinese and English, more functions of background are used in physical sciences than in social sciences; moreover, Chinese writers use more functions background than native speaker of English writers. In terms of critical function, it is obvious that native speaker of English writers employ more critical functions than Chinese writers do; furthermore, in both languages, more critical reporting sentences are found in social sciences than in physical sciences. Sentences expressing critical views in English social sciences occupy the largest percentage while ones in Chinese physical sciences a smallest percentage.

Table 4.24 Raw counts of reporting functions according to language and area

Functions	Chinese social sciences	Chinese physical sciences	English social sciences	English physical sciences
	Raw count	Raw count	Raw count	Raw count
Background	500	482	448	540
Support	21	6	32	13
Critical	14	1	70	38

A two-way ANOVA was employed to examine whether the effect of language (Chinese or English) and area (social and physical sciences) show significance in the difference of choice of reporting functions. The statistic software Matlab 2007b is used and the function of two-way ANOVA is *P=anova2(X)*. *X* in the function *P=anova2(X)* is the corresponding matrix.

There are two variables in this ANOVA: language and area. The data consist of a matrix in the form of 2*2.The variable language includes two levels: Chinese and English, which are the two columns of the matrix. The variable area includes two levels: social sciences and physical sciences, which are the two rows in the matrix. P value can be obtained by using the tool Matlab 2007b. P value is a vector and it has two probability values. If the former $p < 0.05$ or $p < 0.01$, that is to say, the effect of language shows significance. If the latter $p < 0.05$ or $p < 0.01$, that is to say, the effect of area shows significance.

Table 4.25 Variables in two-way ANOVA

Variables	Level 1	Level 2
Language	Chinese	English
Area	Social sciences	Physical sciences

Three reporting functions are examined by two-way ANOVA separately. First of all, the function of background is analyzed. The following is analyzing process:

1) Analysis for function of giving background

Table 4.26 Data of function of background

Variable	Chinese	English
Social	500	448
Physical	482	540

Enter these data:

```
>>a = [500,448;482,540]
```

Obtain this matrix:

```
a =

    500    448
    482    540
```

Enter the function:

```
>> p=anova2(a)
```

Obtain the result:

```
p =

    0.9653    0.6230
```

Table 4.27 ANOVA table 1

Source	SS	Dr	MS	F	Prob>F
Columns	9	1	9	0	0.9653
Rows	1369	1	1369	0.45	0.623
Error	3025	1	3025		
Total	4403	3			

Table 4.27 shows a detailed result for analysis but in this study only the value *Prob>F* is needed. P has two values 0.9653 and 0.6230, it shows that the effect of language and area have no significant difference in choosing function of background. The same method is applied to analysis of the other two functions.

2) Analysis for function of support

```
>> b=[21,6;32,13]

b =
      21       6
      32      13

>> p=anova2(b)

p =

    0.0746    0.1392
```

Table 4.28 ANOVA table 2

Source	SS	Dr	MS	F	Prob>F
Columns	289	1	289	72.25	0.0746
Rows	81	1	81	20.25	0.1392
Error	4	1	4		
Total	374	3			

It is found that p has two values 0.0746 and 0.1392, it proves that the effect of language and area have no significance to difference of choosing function of support.

3) Analysis for critical function

```
>> c=[14,1;70,38]

c =

     14     1
     70    38

>> p=anova2(c)

p =

    0.2543    0.1283
```

Table 4.29 ANOVA table 3

Source	SS	Dr	MS	F	Prob>F
Columns	506.25	1	506.25	5.61	0.2543
Rows	2162.25	1	2162.25	23.96	0.1283
Error	90.25	1	90.25		
Total	2758.75	3			

Table 4.29shows p has two values 0.2543 and 0.1283, it proves that the effect of language and area have no significance to difference of choosing critical function.

Table 4.30 Comparison of p values in three functions

Functions	P value 1	P value 2
Background	0.9653	0.6230
Support	0.0746	0.1392
Critical	0.2543	0.1283

In summary, in three reporting functions of background, support and critical, the effect of language and area is found to show no significance to the differences between two languages and areas. However, Table 4.30 shows that the effect of

language and area in functions of support and critical has more influence than that in function of background. The possible reason is that giving background information is a basic function of reporting in both Chinese and English, and in both social sciences and physical sciences.

Last, proportion of each function to the total is examined.

Table 4.31 Proportion of each function to all the functions

Functions	Raw counts	Proportion
Background	1970	91.0%
Support	72	3.3%
Critical	123	5.7%
Total	2165	100%

Most reporting statements overwhelmingly function as background. There are 5.7% supportive statements and 3.3% statements express critical views. This is consistent with the results of comparison between Chinese corpus and English corpus, between social and physical sciences.

4.5 Summary of main differences between Chinese and English data

To sum up, differences have been found in Chinese and English data in terms of reporting structure, reporting forms, use of reporting verbs and functions of reporting. It is found that Chinese writers use fewer integral citations than native speakers of English. Besides, more non-integral citations are also found in English corpus. Integral and non-integral citations are closely associated with prominence. Integral citations give prominence to cited authors while in non-integral citations information is given prominence. It is concluded that more author prominent citations and author & information citations are employed in English corpus than in Chinese corpus. More citations are also found in English corpus in terms of information prominent citations, weak author prominent citations and general statements.

In terms of reporting forms, it is found that, in both Chinese corpus and English

corpus, summary shares a highest percentage, with generalization following the second. In categories of short direct quotes, extensive direct quotes, summary, generalization and list for reference, there are more counts of reporting statements in English corpus than in Chinese corpus. The biggest difference between two corpora is found in list for reference.

Reporting verbs are import signals to identify reporting statements. Categories in terms of denotation and evaluation and tense of reporting verbs were compared in Chinese and English data. As regards denotation of reporting verbs, more reporting verbs were found in English corpus than in Chinese on the whole. More textual and research verbs are used in English corpus while more mental verbs are used in Chinese corpus. Comparing and theorizing verbs are rarely identified in either of the two corpora. As for evaluative potential of reporting verbs, it is found negative reporting verbs in author's stance and counter-factive ones in writer's stance are used least in both corpora. There is about one time positive and non-factive verbs in English corpus than in Chinese corpus. Chinese has a totally different system to express tense from English. Present, past tense and present perfect are found in English while a small part of past tense can be identified in Chinese; however, the rest can be considered as present or past tense, and past tense or present perfect, which depends.

4.6 Summary of main differences between science and social science data

More integral citations and less non-integral citations are found in social sciences than in physical sciences. Writers in social sciences tend to give prominence to the cited author while writers tend to emphasize information. Besides, writers in social sciences use more repeated names of previously mentioned researcher without citations.

Summary is used most frequently in both social sciences and physical sciences. More short direct quotes, extensive direct quotes, summaries and list for reference are found in social sciences than in physical sciences while more generalizations are found in physical sciences than in social sciences.

More textual and research reporting verbs are identified in social sciences than in physical sciences. Rare comparing and theorizing verb is found in both corpora and only three theorizing verbs are found in social sciences. As to evaluative potential of reporting verbs, more counts of each category in author's stance and writer's stance are found in social sciences than in physical sciences. It is also found that positive and non-factive verbs are used most frequently in both corpora in terms of author's stance and writer's stance respectively. Negative and counter-facitve verbs are least used in both corpora.

In analysis of functions of reporting, the results show that a large number of reporting statements are giving background information while a smallest percentage of reporting statements are providing support to an argument, with the rest expressing critical views. More normed counts of each function are found in social sciences than in physical sciences.

Social sciences which depends on the exercise of individual interpretation is generally less amenable to collaborative publication than physical science which rest on commonly defined criteria of vality (Becher 1989 p.102).

Chapter 5 Conclusion

5.1 General findings

This study examined similarities and differences of reporting between Chinese theses by Chinese writers and English theses by native speakers from four aspects: integral-ness and prominence, reporting forms, reporting verbs and functions of reporting. On the whole, English writers use more reporting statements than Chinese writers. English writers use more integral and non-integral citations than Chinese writers; however, the difference of non-integral citations is more distinct. In total citations of both Chinese data and English data, non-integral citations are used more frequently than integral citations. That is to say, there are more information prominent statements in both corpora. In proportion of integral-ness according to language and area, English physical sciences use most non-integral citations and fewest integral citations while Chinese social sciences use most integral citations and fewest non-integral citations.

As to reporting forms, summary and generalization are used most frequently in both Chinese and English corpora. In category of direct quotes, summaries, generalizations and list for reference, there are more normed counts in English corpus than in Chinese corpus. In proportion according to language and area, summary is used most frequently in four fields and short direct quotes are least used. In total reporting forms in both Chinese and English data, summary shares a highest percentage and generalization follows the second.

It is found that textual verbs and research verbs occur more frequently in English corpus than in Chinese corpus; however, mental verbs are used more frequently in Chinese corpus. Comparing and theorizing verbs are hardly found in either of the two corpora. In total reporting verbs in Chinese data and English data, textual verbs are used most frequently and research verbs rank the second highest percentage. In each category of evaluation, more normed counts are found in English corpus than in Chinese corpus. In neither of the two corpora negative and counter-factive verbs are hardly found. Factive and positive verbs are used most frequently in both corpora. In total reporting verbs, non-factive verbs occupy a highest percentage in writer's stance and positive verbs share a highest percentage in author's stance.

In regard to tense of reporting verbs, past, present tense and present perfect are identified in English corpus, among which, past and present tense are most used with almost the same number. In contrast with English, Chinese is more complex. Expression of tense of most reporting statements in Chinese is not definitely judged, which can indicate present time or past time, past time or present perfect.

Three functions are identified in sample: background, support and critical. In each function, more normed counts are found in English corpus than in Chinese corpus. In both Chinese and English corpora, function of background is used most frequently and supporting function is least used. It is the same case with proportion of function to total number and four fields according to language and area. In three reporting functions, a two-way ANOVA shows the effects of language and area have no significance to the differences between two languages and areas.

5.2 Limitations of this study

Although this study provided a detailed quantitative analysis of reporting practices in master's theses, the limitations of this study are clear: firstly, this study focuses on some aspects of reporting, some other important aspects such as reporting adjuncts (including reporting adverbs, prepositional phrases, subordinative finite clauses), reporting adjectives, and reporting nouns etc. are not included in this paper. Secondly, the English theses are randomly chosen from Internet according to the authors' names. It cannot ensure that the English theses are written by authentic native speakers of English. Lastly, this study is restricted to analysis of master's theses; therefore I'm not sure whether the findings of this study can apply to other genres of academic writing or not.

5.3 Suggestions for further research

This study compares reporting between Chinese theses and English theses, and there are some issues that need further investigation and exploration. Many studies on

English academic writing of Chinese learners of English and contrastive studies on CLE and NS of English indicate CLE use reporting differently from NS. Chinese L2 writers have received intensive criticisms for having a lot of problems in reporting (including citation, plagiarism) in academic writing. The present study can be used as a basis to explore 'why Chinese learners of English behave differently from native speakers of English in reporting in English'. The reasons for reporting practices of Chinese learners of English in constructing English academic writing may be investigated from cognitive perspectives, cross-cultural factors, or language transfer etc. This study may promote people to reconsider these questions: Is it Chinese L2 writers' problem if there is a far distance for them to behave like a native speaker of English in English academic writing? Is it reasonable to expect Chinese learners of English to emulate native model of English academic writing? In a nutshell, the present study is a preliminary investigation, and more comprehensive investigations and further explorations are needed.

References

Becher, T., 1989. *Academic tribes and territories: intellectual enquiry and the cultures of disciplines.* Milton Keynes: Society for Research into Higher Education & Open University Press.

Becher, T., 1989. *Academic tribes and territories: Intellectual inquiry and the cultures of cisciplines.* Milton Keynes, SRHE/Open University Press.

Benson, C., Gollin, J. & Trappes-Lomax, H., 2005. 'Reporting strategies in academic writing: from corpus to materials'. In O. Alexander ed. 2007., *Proceedings of BALEAP Conference 2005.*

Bevalas, J. B., 1978. The social psychology of citations. *Canadian Psychological Review,* 19, p.158-63.

Bloch, J. G., 1990. Universalism versus particularism: The relationship between Chinese and English rhetoric. Paper presented at the 1990 Conference on College Composition and Communication Convention, Chicago, IL.

Bloch, J. G. & Chi, L., 1995. A comparison of the use of citations in Chinese and English academic discourse. In Belcher & Braine eds. *Academic Writing in a Second Language*: *Essays on Research & Pedagogy.* Norwood, NJ: Ablex Publishing Corporation.

Bruce, N. J., 1989. The roles of analysis and the conceptual matrix in a process approach to teaching academic study and communication skills. In V. Bickley, ed. *Language Teaching and learning Styles within and across Cultures*. Hong Kong: Hong Kong Education Press.

Dubois, B. L., 1988. Citation in Biomedical Journal Articles. *English for Specific Purposes,* 7, p.181-193.

Een, J. A., 1982. 'Tense usage in the reporting of past research in geotechnical writing.' *Minnesota Working Papers in ESL*. Minneapolis, Minn.: University of Minnesata.

Gilbert, N. G., 1977. Referencing as persuasion. *Social Studies of Science,* 7, p. 113-22

Granger, S., 1993. 'New snsights into the learner lexicon: a preliminary report from the International Corpus of English'. In L. Flowerdew, & K. Tong, eds. *Entering text*. Hong Kong: The Hong Kong University of Science and Technology

Gosden, H., 1993. Discourse functions of subject in scientific research articles. *Applied Linguistics,* 1, p.56-75.

Groom, N., 2000. Attribution and averral revisited: Three perspectives on manifest intertextuality in academic writing. In P. Thompson, ed. *Patterns and perspectives: Insights for EAP writing practice*. Reading, UK: CALS, The University of Reading, p.15-26.

Hanania, E. & Akhtar, K., 1985. Verb forms and rhetorical function in science writing: A study of MS theses in Biology, Chemistry and Physics. *ESP Journal,* 4(1), p.45-58.

Hyland, K., 1999. Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics,* 20(3), p.341-67.

Hu, Z. & Jiang, Y., 2007. A contrastive study on reporting verbs in English M.A. Theses. *Studies in Languages and Linguistics,* 27(3), p.123-126.

He, C. & Zhou, J., 2001. A Corpus-based study on reporting verbs in English academic writings. *Foreign Language Research,* 107(4), p.43-48.

Jacoby, S., 1987. References to other researchers in literary research articles. *ELR Journal,* 1, p. 33-75.

Kolb, D. A., 1981. 'Learning styles and disciplinary differences' In A. Chickering, ed. *The Modern American College.* San Francisco: Jossey Bass. p.232-55.

Lackstrom, J. E., Selinker, L., & Trimble, L., 1970. Grammar and technical English. In R. C. Lugton ed. *English as a second language." Current issues.* Philadelphia: Center for Curriculum Development.

Lin, H. T., 1984. *Essential grammar for modern Chinese.* Boston, Mass.: Cheng & Tsui Company.

Malcolm, L., 1987. What rules govern tense usage in scientific articles? *English for Specific Purposes,* 6, p.31-44.

Myers, G., 1990. *Writing Biology: Texts in the social construction of scientific knowledge.* Madison: University of Wisconsin Press.

Odlin, T., 1989. *Language transfer: Cross-linguistic influence in language learning.* Cambridge: Cambridge University Press.

Oster, S., 1981. The use of tenses in reporting past literature. In Selinker, Tarone, & Hanzeli, eds. *English for academic and technical purposes: Studies in honour of Louis Trimble.* Newburgh House.

Pickard, V., 1995. Citing previous writers: what can we say instead of 'say'? *Hongkong Papers in Linguistics and Language Teaching,* 18, p.89-102.

Shaw, P., 1992. Reasons for the correlation of voice, tense, and sentence function in reporting verbs. *Applied Linguistics*, 13(3), p.302-19.

Swales, J., 1981. *Aspects of articles introductions.* Birmingham: University of Aston

Swales, J., 1986. Citation analysis and discourse analysis. *Applied Linguistics*, 7(1), p.39-56.

Swales, J., 1990. *Genre analysis.* Cambridge: Cambridge University Press.

Swales, J. & Feak, C. B., 2004. *Academic writing for graduate students: Essential Taskd and Skills.* 2nd ed.. Ann Arbor: The University of Michigan Press.

Sakita, T. I., 2002. *Reporting discourse, tense, and cognition Amsterdam.* London : Elsevier

Thomas, S. & Hawes, T., 1994. Reporting verbs in medical journal articles. *English for Specific Purposes*, 13(2), p.129-48.

Thompson, G. & Ye, Y., 1991. Evaluation in the reporting verbs used in academic papers. *Applied Linguistics,* 12(4), p.365-82.

Thompson, G., 1996. Voices in the text: discourse perspectives on language reports. *Applied Linguistics,* 17, p.501-30.

Thompson, P., 2000. Citation practices in Ph. D. theses. In L. Burnard & T. McEnery, eds. *Rethinking language pedagogy from a corpus perspective: papers from the Third International Conference on Teaching and Language Corpora* . Frankfurt: Peter Lang. p.91-101.

Thompson, P. & Tribble, C., 2001. Looking at citations: using corpus in English for academic purpose. *Language Learning & Technology*, 5 (3), p.91-105.

Weissberg, R. & Buker, S., 1990. *Writing up research: experimental research report writing for students of English.* Prentice Hall.

Yip, Po-Ching. and D, Rimmington., 1998. *Basic Chinese: A Grammar and Workbook*. London : Routledge.

Appendix I Details of Two corpora

Table 1 Details of Chinese data

No.	Author	Year	Title	University
1	Su Mingxia	2005	Studies on University Students' Purchase Decision Strategies and Influence Factors	Nanjing Normal University
2	Zhang Na.	2006	The Impact of Emotional Intelligence on Career Decision-making Process	East China Normal University
3	Li Chao	2005	The Status Quo of the Social Support Seeking of The College Student and the Inquiry of Its Influence Factors	Beijing Normal University
4	Fang Fengjuan	2006	A Research on Infant Mental Health. East China Normal University	East China Normal University
5	Gao Xiuping	2006	A Study on the Develop of Primary School Children's Theory of Mind and its Relation to Aggressive Behavior	East China Normal University
6	Chen Qingchu	2007	Husserl: Phantasy and Pure Possibility	Peking University
7	Xiaohong	2007	Radical needs and the revolution for necessary needs--- An analysis of the theory of human need in Heller	Peking University
8	Gao Song	2007	A Treatise on "The Fact of Reason" of Kant's Practical Philosophy	Peking University
9	Wang Yi	2007	A Priori under the Background of Naturalistic Epistemology	Zhejiang University
10	Wen Ya	2006	The Basic Academic Explanations of Sen's Freedom View	Beijing Normal University
11	Lin Guojin	2007	The British Transport Revolution and Its Influence between 1750-1830	Zhejiang University
12	Lin Ke	2007	The Rise of French Christian Democracy---from the Enlightenment to First World War	Zhejiang University
13	Wang Fang	2007	On Harold Macmillan and the Suez Canal Crisis	Zhejiang University
14	Chen Shanshan	2007	The Changed Psychological State and the Emergence of Enlightenment Thoughts of the Group of adherents of the Ming Dynasty	Zhejiang University

15	Li Jie	2007	Analysis of Property Rights System of Water Resources in Tang Dynasty	Zhejiang University
16	Wei Cong	2004	A Research on the Health Demand	Zhejiang University
17	Cen Cheng	2002	Health Economics and Medical Insurance System Reform in China	Zhejiang University
18	Wu Zhenhui	2003	Transaction Cost and General Equilibrium	Zhejiang University
19	Chen Feng	2002	Economic Analysis on Water Rights Transfer	Zhejiang University
20	Yao Ruqing	2003	Associating Property with Pre-empt and not--- Economic analysis on utilizing valley water resource	Zhejiang University
21	Liu Yang	2004	Periplasmic Proteins of E. coli are Highly Resistant to Aggregation2	Tsinghua University
22	Zhao Feng	2005	Initial Study on the Role of Reactive Cysteine in Arginine Kinase from Sea Cucumber Stichopus japonicus	Tsinghua University
23	Zhang Wei	2008	Screening and Diversity Revelation of PKS and NRPS gene in Sponge Associated Microbes	Shanghai Jiaotong University
24	Zhang Tian	2008	Study on Expression of “Mini C” Human Proinsulin in Pichia Pastoris	Shanghai Jiaotong University
25	Chen Linlin	2008	Genetic Polymorphism Analysis of Cytochrome P4502C19 in Chinese Han Population	Shanghai Jiaotong University
26	Zhang Zhuoqun	2008	Nonlinear Optical and Photoluminescence Study of Indium Nutride Thin Films	Shanghai Jiaotong University
27	Niu Xiaolong	2008	The Absorptive Spectra of Semiconductor Quantum Dot	Shanghai Jiaotong University
28	Liu Lintao	2007	Theoretical Study of Opacity in Plasmas: Simulation of plasmas’ ionic population beyond the average atom model	Shanghai Jiaotong University
29	Xue Ouchen	2008	Thermoelectric Effect of Quantum Dot System	Shanghai Jiaotong University
30	Chang Han	2005	First-principles Study on Hydrogen Adsorption Induced Metallization of SiC Surface	Tsinghua University

31	Zhang Zhenyu	2006	Design of Implanted Central Nerve Recovery System and Its Stimulation Circuits	Southeast University
32	Wang Minsheng	2005	Study and Improvement of Silicon based Germanium Quantum-dot Near Infrared Photodetecter	Tsinghua University
33	Xiong Jingmin	2005	Design of a Programmable Gain Amplifier Circuit	Tsinghua University
34	Lin Yun	2005	Design of High-Speed Multiplexer	Southeast University
35	Xue Zhaofeng	2006	15-40GB/s High Speed Parallel Front-end Amplifiers for Optical Receiver Design	Southeast University
36	Zhang Yuhui	2003	Study on Novel Synthetic Methods of Molecularly Imprinted Polymers and Their Application in the Molecular Recognition	Tsinghua University
37	Li Enhua	2004	Electrochemical Polymerization of β–Naphthalene Sulfonic Acid in Mixed Electrolytes Containing	Tsinghua University
38	Zhang Zhi	2005	Study on Photoelectric Conversion Characteristics of Cuprum Phthalocyanine/Perylene Diimide Molecular System	Tsinghua University
39	Jiang Yong	2003	Preparation of Microspheric Transaconitic Acid Imprinted Polymer and Its Molecular Recognition Function	Tsinghua University
40	Zheng Hongchao	2005	Synthesis of 3’,5’-Dithio-2’-Deoxynucleosides and Study on Their Polymerization	Tsinghua University

Table 2 Details of English data

No.	Author	Year	Title	University
1	Katarina Jane Bellamy	2005	Using the DRM false memory recall paradigm to investigate hemispheric asymmetry and sex differences	The University of Edinburgh
2	Heather Hampton Devine	2005	Effect of Scroll Bar and Navigation Menu Co-location on Web Performance	San Jose State University
3	Jesse Corre Sarubbi	2005	Demystifying Self-harm Behavior Among Teens and Adolescents: A social constructionist approach	State Universtiy of New York
4	Angela Saghar Farrehi	2005	Unique Effect of Individualism Collectivism on Exposure and Reactivity to Daily Stress	University of Delaware
5	Susannah Moat	2005	Phonological Similarity and Lexical Bias in Phonological Speech Errors: self-monitoring or feedback?	The University of Edinburgh
6	Chris Allen	2006	Talking to Both Sides	The University of Edinburgh
7	D. A. Holiday	2005	“Determinacy of sense” in Frege and Wittgenstein	The University of Edinburgh
8	Margot Freedman	2004	Paternalistic Tort Law	Binghamton University
9	Frederic Morneau	2002	Rawls on Global Distributional Justice	Concordia University
10	Daniel Bruce Hansen	2003	A Comparative Account of Substantive Technology Theory in the 20th Century	Lakehead University
11	Mason, Bobby Ray, Jr.	1990	Milovan Djilas and Roy Medvedev: A comparative intellectual history of two dissisents	The University of Texas
12	Misty D. Rodeheaver	2005	An Analysis of the Shifts in Cultural Flows Between the United States and Germany, 1890-1929	West Virginia University
13	Sarah Grehl	2004	The Still Lifes of Francisco De Zurbaran: A departure from the ordinary	California State Universtiy
14	Sarah Thomsen Vierra	2006	Representing Reality: Literature, film, and the construction of Turksih-German Identity	University of North Carolina
15	Heidi Jordan Gardner	2002	The Debate on Luxury in Eighteenth-century Paris: Social stratification and it influence on the arts	University of Missouri-Kansas City

16	David Flores Serrano	2007	The ‘Route Development Fund’. A theoretical model about the likely impact of the Scottish Executive subsidy policy in the behaviour of the Airline Industry	The University of Edinburgh
17	David Dunbar	2004	The Future Management of Biomedical Research in Britain	The University of Edinburgh
18	Simon Kennedy	2004	UK Fixed Line Telecommunications: An assessment of the threat from disruptive technology	The University of Edinburgh
19	John Mayhew	2003	Does the Weather Affect the Financial Performance of Companies in the UK?	The University of Edinburgh
20	Jason Myers	2002	Generic Advertising and the Investment Trust Industry	The University of Edinburgh
21	Richard John Bedoes	1980	New Applications of Grain Shape and Morphological Analysis in Material Science	The University of Calgary
22	Kennedy, Rebecca Ruth, M.S.	1989	Modulation of Delayed-type Hypersensitivity in Mice during Infection with Trichinella pseudospiralis or Trichinella spiralis	The University of Texas
23	Ghazizadeh, Soosan, M.S.	1988	Modulation of cell-mediated Immunity in Mice Infected with Trichinella Pseudospiralis	The University of Texas
24	Maxwell Nicholas Burton	2007	The Mating and Reproductive Behaviour of the Gregarious Parasitoid Wasp, Nasonia vitripennis	The University of Edinburgh
25	Katharine Emma Brooks	2003	Uptake and Presentation of Antigen by B Cells	The University of Edinburgh
26	Bolla, Srinivas Yadav Kumar, M.S.	1994	Finite Element Analysis and Experimental Validation of Ultrasonic Assisted Removal of Colloidal Particles in Fluids	Mississippi State University
27	Mary Hayden Hovater	2005	Hypervelocity Impact Investigation of Solar Sail Materials	The University of Alabama
28	Karen L. M. Nelson	2004	Lidar Observation of Oscillations in the Middle Atmosphere	Utah State University

29	Charles Edward Schroeder	2004	Lag Measurement in an a-Se Active Matrix flat-panel Imager	University of Alberta
30	Jesus Escamilla Roa, B.S.	2003	A Model of Nuclear Matter	The University of Texas
31	Michael M. Boulos	2003	Initial Investigation of the Electronic Structure of Propylene Using Magnetic Circular Dichroism with Synchrotron Radiation in the Vacuum Ultraviolet Region	Florida Atlantic University
32	Jerry Hubert Ross	1995	A Methodology for System Engineering a Medium Scale System and its Application to the Geophysical Airborne Survey System	Mississippi State University
33	Elizabeth Miller	2006	Reduce Cycle Time for Distribution of International Registration Dossiers(IRD): Electronic System	Northern Kentucky University
34	Tinoosh Mohsenin	2004	Design and Evaluation of FPGA-Based Gigabit-Ethernet/PCI Network Interface Card	Rice University
35	Jonathan Sewter	2005	Electronic Equalization of Polarization-mode Dispersion in 40-GB/S Optical Systems	University of Toronto
36	Robert Lester Murry	1996	Continuum Electrostatic Analysis of DNA Bending	Massachusetts Institute of Technology
37	David James Krodel	2000	Photochemical and Time-Resolved Spectroscopic Studies of Dirhodium Hydrogen-Generating Species	Massachusetts Institute of Technology
38	Takashi Nakai	2007	Asymmetric Synthesis of Amines by the Catalytic Enantioselective Additions of Hydrazoic Acid to Ketenes	Massachusetts Institute of Technology
39	Carrie Preston Jones	2007	Development of a Copper-catalyzed Amidation-base-promoted Cyclization Sequence for the Synthesis of 2-Aryl-and 2-Vinyl-4-quinolones	Massachusetts Institute of Technology
40	Ian P. Silverwood	2002	Synthesis and Characterisation of Au/Fe2O3 Catalysts	The University of Edinburgh

Appendix II Details about word counts

Table 1 Details of word counts in Chinese corpus and English corpus

Disciplines	Chinese corpus	English corpus
Psychology	36900	10500
Philosophy	11000	12310
Economics	35400	12000
History	18150	15930
Biology	51000	23200
Physics	12100	18330
Electronic engineering	23000	10500
Chemistry	40870	15250
Average count	5711	2951
Total count	228420	118020

Table 2 Details of word counts in social sciences and physical sciences

Areas	Disciplines	Word counts
Social sciences	Psychology	47400
	Philosophy	23310
	Economics	47400
	History	34080
	Total	152190
Physical sciences	Biology	74200
	Physics	30430
	Electronic engineering	33500
	Chemistry	56120
	Total	194250